UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE AUTOMAÇÃO E SISTEMAS

Feres Azevedo Salem

A model for simulating spreading processes based on social interactions in complex networks: Case studies on online social networks and epidemics

Florianópolis
2023

Feres Azevedo Salem

A model for simulating spreading processes based on social interactions in complex networks: Case studies on online social networks and epidemics

Tese submetida ao Programa de Pós-Graduação em Engenharia de Automação e Sistemas da Universidade Federal de Santa Catarina para a obtenção do título de doutor em Engenharia de Automação e Sistemas.
Orientador: Prof. Ubirajara Franco Moreno, Dr.

Florianópolis
2023

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Salem, Feres Azevedo
A model for simulating spreading processes based on social interactions in complex networks : Case studies on online social networks and epidemics / Feres Azevedo Salem ; orientador, Ubirajara Franco Moreno, 2023.
134 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia de Automação e Sistemas, Florianópolis, 2023.

Inclui referências.

1. Engenharia de Automação e Sistemas. 2. Modelagem de Sistemas. 3. Simulação Computacional. 4. Processos de Espalhamento. 5. Redes Complexas. I. Moreno, Ubirajara Franco. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia de Automação e Sistemas. III. Título.

Feres Azevedo Salem

A model for simulating spreading processes based on social interactions in complex networks: Case studies on online social networks and epidemics

O presente trabalho em nível de doutorado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Carlos Andrey Maia, Dr.
Departamento de Engenharia Elétrica
Universidade Federal de Minas Gerais

Prof. Marcio Viera de Souza, Dr.
Departamento de Engenharia do Conhecimento
Universidade Federal de Santa Catarina

Prof. Eugênio de Bona Castelan Neto, Dr.
Departamento de Automação e Sistemas
Universidade Federal de Santa Catarina

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de doutor em Engenharia de Automação e Sistemas.

ICP-Edu
Documento assinado digitalmente
Julio Elias Normey Rico
Data: 13/10/2023 16:34:13-0300
CPF: ***.840.859-**
Verifique as assinaturas em https://v.ufsc.br

Prof. Julio Elias Normey-Rico, Dr.
Coordenador do Programa

ICP-Edu
Documento assinado digitalmente
Ubirajara Franco Moreno
Data: 13/10/2023 15:06:02-0300
CPF: ***.579.069-**
Verifique as assinaturas em https://v.ufsc.br

Prof. Ubirajara Franco Moreno, Dr.
Orientador

Florianópolis, 2023.

To Isabela, Delfina, Abdul, and all my beloved family.

ACKNOWLEDGEMENTS

Os anos que se passaram durante esse curso de doutorado foram regados de alegrias, mas também de muitas dificuldades tanto para mim quanto para a sociedade no geral. Tantas pessoas deixaram sua marca nesse pedaço da minha história que, para evitar ser injusto ou me esquecer de alguém, vou procurar não citar muitos nomes nesses agradecimentos.

Sou grato à minha família, pelo amor incondicional, por todo o carinho, apoio e paciência que me sustentaram na perseguição desse sonho. Agradeço especialmente à Isabela, pelas tantas vezes que não me deixou desistir durante a escrita dessa tese, por ser a minha companheira, no sentido mais puro da palavra, e por compartilhar comigo os sonhos, as aventuras e a vida.

Sou grato ao Prof. Ubirajara Franco Moreno (ou simplesmente Bira), pelos quase sete anos de orientação e parceria, por exigir bastante de mim, mas sem nunca deixar de demonstrar humanidade nos momentos em que precisei. Agradeço igualmente aos professores e técnicos do DAS/UFSC, que me apoiaram e tanto me ensinaram durante esse período, e que, hoje, felizmente, considero muitos deles como amigos. Não posso deixar de agradecer aos professores que me impactaram antes do mestrado e foram fundamentais, como exemplos e incentivadores, na escolha pela vida acadêmica.

Sou grato aos professores membros da banca pelo tempo dedicado em contribuir para o aperfeiçoamento dessa tese. Agradeço ao relator, Prof. Carlos Andrey Maia, pelos seus importantes comentários e pela forma construtiva com que redigiu o seu parecer sobre meu trabalho.

Sou grato aos amigos, alguns de longa data, outros que fiz durante a pós-graduação, que são a minha família em Floripa. Obrigado por estarem presentes, por ouvirem minhas reclamações e por celebrarem comigo cada vitória. Sou grato pela “salinha dos doutorandos”, pelas pessoas que lá estavam quando cheguei e por aquelas que chegaram depois. Obrigado por cultivarem esse ambiente de troca de ideias, amizade e, muitas vezes, sessões coletivas de terapia.

Sou grato ao Prof. Giuseppe Notarstefano por aceitar a orientação durante o doutorado sanduíche na Universidade de Bolonha, me permitindo vivenciar diferentes culturas e formas de se fazer pesquisa. Obrigado aos amigos que fiz nos laboratórios OPT4SMART e CIRSFID, por tornarem essa experiência única.

Agradeço à UFSC, pela estrutura fornecida, ao CNPq e à CAPES, pelas bolsas de doutorado e de intercâmbio. Sou grato também ao Instituto SENAI de Inovação, pela oportunidade de trabalhar com pesquisa aplicada dentro da minha área de conhecimento e com um time de pessoas competentes e dedicadas.

Por fim, sou grato a Deus, pelas muitas bençãos que recebi ao longo desses anos e que tentei descrever brevemente acima.

“One of the great challenges in life is knowing
enough to think you are right, but not
knowing enough to know you are wrong.”
(Neil deGrasse Tyson)

RESUMO

Esta tese foi motivada por dois problemas relacionados aos processos de espalhamento: a desinformação sendo disseminada em Redes Sociais Online (OSNs) e as doenças infecciosas sendo transmitidas em uma população suscetível. O principal objetivo do presente trabalho é desenvolver um modelo para simular processos de espalhamento baseado em interações sociais e que permita abordar tanto a estrutura complexa das redes, como os comportamentos evolutivos dos agentes em um nível microscópico. Para tanto, é fornecido um modelo genérico para interações sociais assíncronas entre agentes, o qual é estendido tanto para a troca de informações como para os casos de transmissão de doenças. Um modelo de evolução da infecção é construído, dispondo de transições probabilísticas entre os estágios, permitindo o uso de dados do mundo real de forma direta. Esse modelo de troca de informações pode tratar conjuntos finitos de informações, diferentemente dos modelos da literatura de dinâmica de opinião. Também, neste último estudo de caso, outros fatores que impactam a disseminação são considerados, como a autoestima da pessoa e a confiabilidade da informação. Ambos os fatores são modelados como uma extensão do modelo anterior, porém, admitindo matrizes dinâmicas para as probabilidades das interações. Métricas para capturar informações relevantes sobre a estrutura de redes complexas grandes são estudadas, concluindo que a centralidade de autovetor está intimamente relacionada à velocidade de propagação e à probabilidade de uma determinada informação prevalecer sobre as demais no modelo proposto. Estruturas de rede estáticas e dinâmicas são construídas para representar cenários relevantes em ambos os estudos de caso, as quais são baseadas em modelos de redes complexas encontrados na literatura. Além disso, um algoritmo de clusterização é modificado para identificar comunidades em redes sociais. Este algoritmo alterado é capaz de evitar erros que foram identificados na utilização de algoritmos da literatura. Por fim, conhecer como se dá a divisão em comunidades de uma rede social, e alguma métrica de centralidade sobre os agentes, permite que políticas para controlar a disseminação de informações nas OSNs sejam propostas.

Palavras-chave: Processo de Espalhamento. Simulação Baseada em Multiagentes. Redes Complexas. Notícias Falsas. Covid-19. Modelo Probabilístico. Dinâmica de Opinião.

ABSTRACT

This thesis is motivated by two problems related to spreading processes: misinformation being disseminated in Online Social Networks (OSNs), and infectious diseases being transmitted in a susceptible population. The main objective of the present work is developing a model for simulating spreading processes based on social interactions, which address both the networks' complex structure and the evolving behaviors of the agents on a microscopic level. To this purpose, we provide a generic model for asynchronous social interactions between agents, which is extended for both the information exchange and the disease transmission cases. We build a model for infection evolution with probabilistic transitions between stages, allowing the usage of real-world data in a straightforward way. Our information exchange model can handle finite sets of information, differently from the models in the opinion dynamics literature. Also, in this last case study, we considered other factors that impact the spreading, like the person's self-esteem and the information's reliability. We model both factors as an extension from the previous one, by admitting dynamic matrices for the interactions' probabilities. We study metrics for capturing relevant information on the structure of large and complex networks, concluding that eigenvector centrality is intimately related to the spreading speed and the probability of a given information prevailing over the others in our model. We construct static and dynamic network structures representing meaningful scenarios for both case studies, which rely on complex network models from the literature. Further, we present the modification for a clustering algorithm to identify communities in social networks. This amended algorithm can avoid errors in the clustering that we have identified while using algorithms from the literature. Finally, we show that knowing how a social network is divided into communities, and some centrality metrics about the agents, enable us for proposing policies to control the spreading of information in OSNs.

Keywords: Spreading Process. Multi-Agent Based Simulation. Complex Network. Fake News. Covid-19. Probabilistic Model. Opinion Dynamics.

RESUMO EXPANDIDO

UM MODELO DE SIMULAÇÃO PARA PROCESSOS DE ESPALHAMENTO BASEADO EM INTERAÇÕES SOCIAIS EM REDES COMPLEXAS: ESTUDOS DE CASO EM REDES SOCIAIS ONLINE E EPIDEMIAS

Introdução

A motivação inicial desse trabalho se deu por um tema que preocupa a comunidade científica e parte da população: o espalhamento de desinformação em Redes Sociais Online (OSNs, do inglês *Online Social Networks*). Isso se evidencia pelo recente aumento no número de publicações relacionadas ao tema, sendo que mais de 75% dos artigos relacionados a desinformação dos últimos 20 anos foram publicados entre 2018 e 2022. Um dos fatores responsáveis para esse aumento de interesse pelo tema foi a introdução das mídias sociais como um ponto decisivo no resultado de eleições pelo mundo. Sendo que os primeiros casos a soar o alarme para esse problema foram a votação pela saída do Reino Unido da União Europeia (Brexit) e a eleição presidencial dos EUA em 2016, onde houve disseminação massiva de conteúdos sendo, muitos deles, falsos.

Outro campo que traz preocupação nesse sentido é o da saúde, no qual a desinformação acarreta a divulgação de métodos e curas sem respaldo científico. Esse problema se agravou significativamente durante a pandemia de Covid-19. Tratando-se de notícias falsas no geral, estudos já mostram o seu potencial de alterar as visões de mundo das pessoas, fazendo-as enxergar uma realidade distorcida, o que acaba dificultando o diálogo e gerando cada vez mais polarização. Além disso, as pessoas têm suas decisões, sejam acerca de saúde, finanças ou política, influenciadas por esses conteúdos. Outro agravante é o fato de que as notícias falsas tendem a se espalhar mais rápido em relação às verdadeiras, o que dificulta ações para reparar o dano causado por elas.

A comunidade de sistemas e controle também tem se preocupado com esse tema, sendo que o mesmo apareceu com certo destaque nas últimas três edições do Congresso Mundial da Federação Internacional de Controle Automático (IFAC, do inglês *International Federation of Automatic Control*). Na edição de 2017, houve um painel para discutir os desafios de controle para sistemas sociais e o presidente do comitê técnico sobre o impacto social da automação trouxe que um dos desafios particularmente seria o uso de redes sociais para controlar a opinião pública. Ainda nesse ano, os membros de um comitê do IFAC publicaram uma agenda de pesquisa para a comunidade de sistemas e controle e apontaram as redes sociais e tecno-sociais como um tema de alto potencial para pesquisa. Em 2020, em uma edição virtual do Congresso Mundial da IFAC, foi apresentada uma sessão tutorial sobre o tema de Redes Dinâmicas de Influência Social. Já na última edição desse Congresso, ocorrida em julho de 2023, além de sessão específica sobre o tema de controle de sistemas socio-técnicos em rede, houve a apresentação do Roadmap 2030 dos Desafios de Controle em Escala de Sociedade elaborado pela Sociedade de Sistemas de Controle (CSS) da IEEE. Entre outros pontos importantes, nesse documento, é trazida à luz a necessidade de se entender o efeito das redes em sistemas com dinâmicas sociais, como no caso de epidemias e pandemias como a de Covid-19. Também apresentam redes sociais de dinâmica de opinião e tomada de decisão coletiva como um desafio notável na modelagem do comportamento humano.

Objetivos

O objetivo geral desta tese é desenvolver um modelo de simulação para interações sociais ocorrendo em redes complexas de agentes com comportamentos evolutivos.

Com base nesse objetivo, esta pesquisa busca duas contribuições principais: melhorar a qualidade e a extensão das análises sobre processos de espalhamento dependentes de interações sociais, e fornecer políticas para a síntese de uma estrutura de rede capaz de controlar esses processos. Para esses fins, foram definidos os seguintes objetivos específicos:

- Desenvolver um modelo para as interações sociais entre agentes interconectados;
- Modificar um modelo de troca de informações para lidar com conjuntos finitos de informações;
- Propor estruturas de rede que representem os cenários estudados, baseando-se em modelos de redes complexas encontrados na literatura;
- Identificar características relevantes que afetam a disseminação de informações, além da estrutura da rede;
- Construir modelos que permitam matrizes dinâmicas de probabilidade de interação para incorporar as características identificadas;
- Construir um modelo para a disseminação de doenças infecciosas com base em interações sociais;
- Criar um modelo estocástico para a evolução de doenças em cada indivíduo, que possa usar dados do mundo real para ajustar seus parâmetros;
- Simular todos os modelos propostos em cenários significativos para avaliar sua capacidade de expandir as possibilidades de análise;
- Estudar características da estrutura da rede que possam influenciar os processos de espalhamento;
- Adaptar um algoritmo de clusterização para identificar comunidades em redes sociais;
- Usar o conhecimento adquirido com essas características para propor políticas de controle da disseminação.

Metodologia

A abordagem utilizada no desenvolvimento dessa tese partiu do estudo dos modelos de epidemiologia, mas com interesse no seu uso para analisar o espalhamento de informações, visto que essa analogia já era bastante explorada na literatura. Conforme os estudos se aprofundaram, notou-se a potencialidade de um modelo mais microscópico, onde o foco seria nos comportamentos individuais e a estrutura da rede seria um fator chave na descrição da dinâmica de espalhamento. Os modelos de dinâmica de opinião presentes na literatura já tinham uma forte dependência da topologia de rede, porém tratavam as opiniões como valores contínuos que eram sempre alterados pelas interações e isso não modelaria adequadamente o caso de pessoas compartilhando notícias e encaminhando mensagens exatamente da forma como as receberam, que é o caso na maior parte das vezes quando se tratam de redes sociais. Portanto, o primeiro passo foi prover um modelo que lidasse com opiniões ou informações discretas. Em seguida, se deu a generalização do modelo de interações sociais. Assim, obteve-se um modelo dividido entre: a descrição de quando e por que uma interação ocorre; e a descrição dos comportamentos dos agentes durante uma interação.

Além disso, para o primeiro estudo de caso, foram identificados diversos fenômenos que poderiam influenciar no processo de disseminação de informações. Dentre esses, foram incorporados na modelagem o efeito no espalhamento da variação na autoestima dos agentes e do conteúdo das informações. Para esse último, foi implementado um coeficiente de confiabilidade atrelado à informação que está sendo compartilhada, uma vez que apenas as relações interpessoais não seriam suficientes para julgar se uma informação é verdadeira ou não. Posteriormente, numa etapa mais focada na síntese da rede para controlar o espalhamento, foi desenvolvido um algoritmo para a identificação de comunidades em redes sociais. E, utilizando essas comunidades detectadas e conhecimentos vindos da análise de centralidade dos nós da rede, foram indicadas políticas para atenuar ou impulsionar o espalhamento, baseadas na identificação dos nós com maior potencial de disseminação.

O segundo estudo de caso trazido nesse trabalho foi motivado pelo início da Pandemia de Covid-19, onde os esforços foram redirecionados no sentido de tentar contribuir com esse problema urgente pelo qual o mundo estava passando. Então, com base no conhecimento de epidemiologia adquirido nas etapas iniciais da pesquisa e na evolução da modelagem para o espalhamento de informações, observaram-se possíveis vantagens na análise e predição do espalhamento de doenças contagiosas. Assim, foi proposto um modelo com abordagem microscópica, explorando os comportamentos individuais dos agentes, o que permite obter informações mais detalhadas e elaborar estratégias mais refinadas para reduzir o contágio. Para tanto, foi utilizado o mesmo modelo de interações sociais proposto anteriormente, mas apresentado um modelo novo para descrever o que ocorre após a interação ou infecção e para a posterior evolução da doença em cada indivíduo.

O ponto em comum entre essas duas aplicações se dá no fato de serem sistemas complexos, influenciados pelo comportamento dos agentes e suas interações. Ao analisar sistemas complexos, algumas simplificações são necessárias, sendo que as mais comuns na literatura seriam a análise separada dos efeitos da topologia e do comportamento dos nossos. Como exemplo de simplificação da topologia, muitos trabalhos consideram uma rede totalmente conectada. Por outro lado, para explorar redes com estruturas complexas, uma simplificação comum para o comportamento dos nós é modelá-los como um integrador simples. Esse trabalho, porém, sugere um modelo que seja complexo nesses dois aspectos: os comportamentos individuais e a topologia de rede.

Resultados e Discussão

Com base no amplo conjunto de simulações retratadas neste documento, pode-se dizer que os modelos propostos melhoram e ampliam a capacidade de análise quando comparados com os modelos provenientes da literatura. Além disso, o conhecimento fornecido por essas análises aprimoradas pode ser usado como uma ferramenta para sintetizar estruturas de rede, a fim de controlar processos de propagação em sistemas tecnossociais.

Foi fornecido um modelo para interações sociais entre agentes e o mesmo foi ampliado para analisar dois estudos de caso principais: informações sendo trocadas por interações sociais *online* e doenças infectando pessoas por meio de interações sociais no *mundo real*. Foi criado um modelo para a evolução da infecção, que é flexível o bastante para representar diferentes doenças e níveis de detalhamento. Esse modelo admite transições probabilísticas entre os estágios da infecção, podendo se beneficiar de dados estatísticos do mundo real para a sua descrição e, também, ser atualizado a partir de novos dados.

O modelo proposto para a troca de informações lida com conjuntos finitos de informações, o que é uma novidade em relação aos modelos de dinâmica de opinião encontrados na literatura. A propagação no modelo desenvolvido é ditada pelas probabilidades da matriz de interação, a qual está diretamente relacionada à estrutura da rede. No entanto, foram identificadas outras características que afetam a disseminação de informações em uma rede social. Assim, foram modelados também o efeito da confiabilidade das informações e da autoestima dos agentes como fatores que alteram dinamicamente as matrizes de probabilidades de interação.

Dado que OSNs são realmente grandes e complexas, qualquer tentativa de extrair suas características por inspeção visual seria infrutífera. Por isso, foram usadas métricas estatísticas para descrever esse tipo de estrutura. Concluiu-se que a centralidade de autovetor está intimamente relacionada à velocidade de propagação e à probabilidade de uma determinada informação prevalecer sobre as demais no modelo apresentado nessa tese.

Para ambos os estudos de caso, foram fornecidas estruturas de rede hipotéticas para uma variedade de cenários. Modelos de redes complexas, como o modelo Watts-Strogatz para redes *Small-World* (ou de “mundo pequeno”), são abordagens melhores para descrever a topologia por trás de interações sociais humanas do que redes regulares ou totalmente aleatórias. Portanto, esses modelos de redes complexas foram base de muitas das estruturas de rede propostas nesse trabalho para representar a organização de grupos, tanto no mundo real como *online*.

O resultado mais recente foi o desenvolvimento de um algoritmo de clusterização aplicado para identificar comunidades em redes sociais. O algoritmo proposto é capaz de evitar erros na classificação que foram identificados ao usar algoritmos da literatura. Também foi mostrado que o conhecimento sobre essa divisão em comunidades de uma rede social e algumas métricas de centralidade sobre os agentes podem ser usados para controlar a disseminação de informações em OSNs.

Considerações Finais

Conclui-se ser necessário aprimorar as predições dos fenômenos de espalhamento estudados para ser possível traçar estratégias e políticas capazes de controlar esses processos. Os modelos propostos nesse trabalho estendem a capacidade de análise

em relação aos disponíveis na literatura especializada. Além disso, o conhecimento sobre as comunidades presentes em uma rede social e a centralidade dos seus membros serve para determinar ações de mitigação ou impulsionamento da disseminação. Por fim, acredita-se no potencial de melhor abordar temas tão contemporâneos e motivadores por meio de grupos de pesquisa interdisciplinares, capazes de congregar especialistas dos diferentes campos de conhecimento necessários para compreender os fenômenos presentes nos processos estudados ao longo desse trabalho.

Palavras-chave: Processo de Espalhamento. Simulação Baseada em Multiagentes. Redes Complexas. Notícias Falsas. Covid-19. Modelo Probabilístico. Dinâmica de Opinião.

LIST OF FIGURES

Figure 1 – Number of publications per year associated with the term “misinformation” in the Scopus platform. 25
Figure 2 – Number of publications per year that relate misinformation with politics, Covid, and others. 25
Figure 3 – Overview of the contributions of the thesis and how they are connected to each other. 32
Figure 4 – Graphs examples. 35
Figure 5 – Random rewiring procedure, from a regular ring lattice to a random network. 37
Figure 6 – Information of each node evolving in time according to the consensus model. 39
Figure 7 – SEIR model diagram. 45
Figure 8 – Diagram for the model altered to address a Covid-19 epidemic. . . . 46
Figure 9 – Overview of the agents’ evolution, where SA means “Susceptible Agent” and IA means “Infected Agent”. 47
Figure 10 – Illustration for the concept of probabilistic transitions between states, in which the probability varies as time passes since the previous transition. 48
Figure 11 – Step-by-step of the simulation process applied to the illustrative example 49
Figure 12 – State machine for the individuals, where: S $\rightarrow$ Susceptible; E $\rightarrow$ Exposed; I $\rightarrow$ Infected; R $\rightarrow$ Removed. 51
Figure 13 – Transition probability from exposed [E] to infected [I] as a function of the number of days since the transition from susceptible [S]. 52
Figure 14 – Transition probability from infected [I] to removed [R] as a function of the number of days since the transition from exposed [E]. 53
Figure 15 – Graph representing the isolated clusters (classrooms). 54
Figure 16 – Example of graph representing the random interactions between all nodes. 54
Figure 17 – Mean and standard deviation for the spreading in the **20-10-6** scenario. 56
Figure 18 – Stacked bars plot for the average number of persons in each state, for the **20-10-6** scenario. 56
Figure 19 – Stacked bars plot for the average percentage of people that caught the virus, for the **20-10-6** scenario. 57
Figure 20 – Histograms of values for the random variable $R_i(k)$ on different replications of the simulations for scenario **20-10-6** 58
Figure 21 – Mean and standard deviation for the spreading in the **15-7-4** scenario. 58

Figure 22 – Stacked bars plot for the average number of persons in each state, for the **15-7-4** scenario. 59
Figure 23 – Stacked bars plot for the average percentage of people that caught the virus, for the **15-7-4** scenario. 59
Figure 24 – Mean and standard deviation for the spreading in the **10-5-4** scenario. 60
Figure 25 – Stacked bars plot for the average number of persons in each state, for the **10-5-4** scenario. 60
Figure 26 – Stacked bars plot for the average percentage of people that caught the virus, for the **10-5-4** scenario. 61
Figure 27 – State machine for the individuals, where: S → Susceptible; Ns → No symptoms; Ms → Moderate symptoms; Ss → Severe symptoms; H → Hospitalized; Hv → Hospitalized using ventilator; R → Recovered; D → Dead. 63
Figure 28 – Example of a graph at the beginning of the simulation. 65
Figure 29 – QR Codes directing for animations showing the Covid spreading evolution according to our model in two different scenarios. 65
Figure 30 – Evolution of cases averaged over 100 simulations, not considering the health system overload. 66
Figure 31 – Evolution of cases averaged over 100 simulations, considering the health system overload. 67
Figure 32 – Comparison of ventilators demanded versus available (averaged over 100 simulations). 67
Figure 33 – Evolution of cases averaged over 1000 simulations. 68
Figure 34 – QR Code directing for the animation illustrating the consensus with information from a finite set, also accessible through the link: https://youtu.be/R3nEx1tqL 72
Figure 35 – Self-weights evolution with 10 agents (humans). 78
Figure 36 – Self-weights evolution with 8 human agents and 2 robot agents. 79
Figure 37 – Evolution of the influence network across the sequence of issues. 79
Figure 38 – Opinion evolution with 8 human agents and 2 robot agents. 80
Figure 39 – Opinion evolution with 90 human agents and 10 robot agents. 80
Figure 40 – Initial conditions for each scenario. 82
Figure 41 – Example of final states in the scenario with one malicious node per group. 84
Figure 42 – QR Code directing for the animation exemplifying the simulation in the scenario depicted in Figure 41a, also accessible through the link: https://youtu.be/0XXX9Q_jeNk. 84
Figure 43 – Example of final states in the scenario with two malicious node per group. 84

Figure 44 – Example of final states in the scenario with three malicious node per group. 85
Figure 45 – Topologies: (a) Linear graph (b) Circular graph with even number of nodes (c) Arbitrary graph (d) Full connected graph (e) Circular digraph with clockwise direction (f) Circular digraph with source (g) Circular digraph with sink (h) Five nodes, each with two connections (i) Six nodes, each with three connections (j) Eight nodes, each with two connections (k) Circular digraph with odd number of nodes (l) Star graph. 89
Figure 46 – Comparison between the occurrence ratio of each initial opinion and the centrality score (red crosses) of the agents in the respective topologies from Figure 45. 90
Figure 47 – Adopted topology. Dashed circles represent an estimation of Kilobots communication range, continuous lines are logical connections between them, and the edge weights represent the importance each robot attributes to its neighbors. 92
Figure 48 – Absolute errors between the eigenvector centrality scores and the aggregate outcomes from all three tests. 93
Figure 49 – Example of applying the 2-Hops Neighborhood Correction Stage to the output of the FCM Spectral Clustering. 103
Figure 50 – Initial condition of the information spreading process 103
Figure 51 – States of the information spreading on the network at sample k. . . . 104
Figure 52 – QR Code directing for the animation illustrating the spreading evolution, also accessible through the link: https://youtu.be/7nFPHSfgzAw. 105
Figure 53 – Output of the K-means spectral clustering, type 1 error present at node 31 106
Figure 54 – Output of the K-means spectral clustering, presenting type 2 error at nodes 10 and 12 106
Figure 55 – Zooming into nodes 10 and 12 makes clearer that they are not connected to any other nodes from their cluster 107
Figure 56 – Output of the FCM spectral clustering, not presenting the same error from K-means 107
Figure 57 – Output of the FCM spectral clustering, the type 2 error persists at nodes 10 and 12 108
Figure 58 – Output of the FCM-2Hops spectral clustering, not presenting the same error as in the previous approaches 108

LIST OF TABLES

Table 1 – Statistical analysis for scenario **20-10-6** 57
Table 2 – Statistical analysis for scenario **15-7-4** 60
Table 3 – Statistical analysis for scenario **10-5-4** 61
Table 4 – Consensus outputs. 78
Table 5 – Results averaged over 1000 simulations for each scenario, considering the blue information as malicious and no information with reliability stamp. 83
Table 6 – Results averaged over 1000 simulations for each scenario, considering the blue information as malicious and the red information as marked by the reliability stamp. 83
Table 7 – Results from the experiments with kilobots. 92
Table 8 – Number of simulations presenting errors of any type in the clustering output for each algorithm 109
Table 9 – Number of samples until the information reaches all nodes (averaged over 1000 simulations), considering 4 clusters and varying the network size. Comparison between the different policies for choosing the spreaders. 110
Table 10 – Number of samples until the information reaches all nodes (averaged over 1000 simulations), fixing the network size as 500 nodes and varying the number of clusters/spreaders. Comparison between the different policies for choosing the spreaders. 110

LIST OF ABBREVIATIONS AND ACRONYMS

kNN	k Nearest Neighbors
AI	Artificial Intelligence
CBA	Congresso Brasileiro de Automática
Covid-19	CoronaVirus Disease 2019
FCM	Fuzzy c-means
FCM-2Hops	FCM Spectral Clustering with 2 Hops Neighborhood Correction
GARCH	Generalized Autoregressive Conditional Heteroscedasticity
ICU	Intensive Care Unit
IEEE	Institute of Electrical and Electronic Engineers
IFAC	International Federation of Automatic Control
IFR	Infection Fatality Rate
NPI	Non-Pharmaceutical Intervention
OCEAN	Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism
OSN	Online Social Network
RS	Recommender System
SARS-CoV-2	Severe Acute Respiratory Syndrome CoronaVirus-2
SBAI	Simpósio Brasileiro de Automação Inteligente
SEIR	Susceptible-Exposed-Infected-Removed
SI	Susceptible-Infected
SIR	Susceptible-Infected-Recovered
SIS	Susceptible-Infected-Susceptible
TC	Technical Committee
UFSC	Federal University of Santa Catarina
USA	United States of America
VMD	Variational Mode Decomposition

LIST OF SYMBOLS

$\mathcal{G}$	Undirected Graph
$\mathcal{N}$	Set of Nodes or Vertices of a Graph
$\mathcal{E}$	Set of Edges or Links of a Graph
$\subseteq$	“Is a subset of”
$\times$	Cartesian product
$\mathcal{A}$	Adjacency matrix of a Graph
α_{ij}	Element in the i-th line and j-th column of $\mathcal{A}$
$\mathcal{D}$	Directed Graph
$\lvert \cdot \rvert$	Cardinality or number of elements of a set
$d(i)$	Degree of node i
$d_{in}(i)$	In-degree of node i
$d_{out}(i)$	Out-degree of node i
$\mathcal{N}_i$	Set of Nodes connected to (or in the neighborhood of) node i
$\in$	“Is an element of” or “belongs to” or “in”
$\vert$	In the context of intensional set definition it means “such that”
$\dot{g}_t$	First derivative of function g_t with respect to time
x_t^i	Vector of information states of node i at continuous time t
x_k^i	Vector of information states of node i at discrete time k
π_{ij}	Weight between nodes i and j in the discrete-time consensus algorithm
$\forall$	“For all” or “for any”
z_k	Vector of opinions at sample k in the DeGroot model for opinion dynamics
$\mathcal{W}$	Row-stochastic matrix for the influence weights
w_{ij}	Element in the i-th line and j-th column of $\mathcal{W}$ representing how much influence j has over i
i_k	Active agent at sample k in the gossip-based consensus algorithm
$\mathcal{P}$	Probability of interaction matrix
p_{ij}	Element in the i-th line and j-th column of $\mathcal{P}$ representing how likely it is for j to interact with i
$\triangleq$	“Is defined to be equal to”
β	Rate of Susceptible individuals becoming Exposed in the SEIR model
ε	Rate of Exposed individuals becoming Infected in the SEIR model
γ	Recovery rate in the SEIR model
δ	Mortality rate in the SEIR model
ς	Average number of contacts for each individual in the SEIR model
S_t	Number of Susceptible individuals at time t in the SEIR model
E_t	Number of Exposed individuals at time t in the SEIR model
I_t	Number of Infected individuals at time t in the SEIR model

R_t	Number of Removed individuals at time t in the SEIR model
ν^i	Constant representing the trust of agent i in its neighbors in the gossip-based consensus algorithm
$\mathcal{W}_s$	Issue-dependent matrix for the influence weights at issue s
$x_{s,k}$	Vector of information for all agents at sample k of issue s
χ_s	Vector of self influence weights on issue s
χ_s^i	Self-weight of node i on issue s
$\mathcal{C}$	Static and issue independent relative interpersonal weights
c_{ij}	Element in the i-th line and j-th column of $\mathcal{C}$
Δ_n	The n-simplex
$\mathbb{e}_i$	The i-th basis vector
σ	Vector of eigenvector centrality scores
σ^i	Eigenvector centrality score for node i
$\mathcal{F}_s$	Confidence range matrix on issue s
$f_{ij}(s)$	Element in the i-th line and j-th column of $\mathcal{F}_s$
φ	Normalized random vector determining the occurence of interactions
$\exists$	“There exists” or “there is at least one”
$\langle \cdot, \cdot \rangle$	Two positions tuple
X_k^i	Information tuple of agent i at sample k
$X_k^{i\langle q \rangle}$	The q-th element of the information tuple that agent i is holding at sample k
τ_m	Numerical measurement of the reliability of information m
$\mathcal{A}_k$	Time-variant adjacency matrix of a Graph at sample k
$\mathcal{P}_k$	Time-variant probability of interaction matrix at sample k
Λ	Vector containing the eigenvalues of $\mathcal{W}$ in an ascending order
λ_i	The eigenvalue of $\mathcal{W}$ with the i-th smallest absolute value
v_i	The left eigenvector associated with λ_i
v_{max}	The dominant left eigenvector of $\mathcal{W}$
$\lVert \cdot \rVert_1$	The 1-norm of a given vector
(a,b)	Open interval of values from a to b
ϑ	Vector containing the occurrence of consensus on each initial information
$\{\cdot, \ldots, \cdot\}$	Extensional definition of a set
θ_n	The n-th observation or data point in a data set
r_{nk}	Binary variable indicating whether θ_n is assigned to cluster k
μ_k	Centroid of the cluster k
J	Objective function for the K-means clustering algorithm
$\lVert \cdot \rVert$	The Euclidean norm of a given vector
$[a,b]$	Closed interval of values from a to b
ρ_{nk}	Membership degree of θ_n to cluster k
J_m	Objective function for the Fuzzy c-means clustering algorithm

m	Constant controlling how fuzzy are the boundaries between clusters in the FCM approach
$\mathrm{diag}(\cdot)$	A diagonal matrix formed by the elements of a vector
D	Degree matrix with the node's degrees in its diagonal
$\mathcal{L}$	Laplacian matrix of a Graph
u_j	The j-th eigenvector of the Laplacian
$\mathcal{U}$	Matrix composed by the K first eigenvectors of the Laplacian as it columns
y_i	The i-th row of $\mathcal{U}$
C_k	The k-th cluster of the points $\{y_1 \dots y_N\}$
$\mathcal{V}_k$	The k-th cluster of the nodes of the graph
$\mathcal{N}_{ik}$	Set of neighbors of i in cluster k
$\wedge$	Logical operator AND
$\tilde{C}_k$	Cluster k identified by the FCM-2Hops approach

CONTENTS

1 INTRODUCTION 24
1.1 PROBLEM, CONTEXT, AND MOTIVATION 24
1.2 PROPOSAL AND NOVELTY 28
1.3 OBJECTIVES 29
1.3.1 Specific Objectives 29
1.4 METHODOLOGY 29
1.5 CONTRIBUTIONS OF THE THESIS 31
1.6 DOCUMENT STRUCTURE 33
2 MODELING SOCIAL INTERACTIONS 34
2.1 MODELING NETWORKS 35
2.2 COMPLEX NETWORKS 36
2.3 CONSENSUS PROBLEM 38
2.3.1 Opinion Dynamics 39
2.3.2 Gossip-based Consensus 39
2.4 MODEL FOR SOCIAL INTERACTIONS 40
3 DISEASES SPREADING 41
3.1 BACKGROUNDS ON MODELING EPIDEMICS 44
3.2 SIMULATION MODEL 46
3.2.1 Infection Evolution Model 47
3.2.2 Simulation Steps 48
3.3 ASSESSING THE RESUMPTION OF PRESENTIAL TEACHING ACTIVITIES 50
3.3.1 Model for the infection evolution 51
3.3.2 Model for the social interactions 53
3.3.3 Simulations and Discussion 55
3.4 PREDICTING THE VIRUS SPREAD IN A CITY 62
3.4.1 Model for the infection evolution 62
3.4.2 Model for the social interactions 64
3.4.3 Simulations and Discussion 66
4 INFORMATION SPREADING 69
4.1 INFORMATION EXCHANGE MODEL 71
4.1.1 Modelling the Effect of Self-esteem 73
4.1.2 Simulations for Humans Self-esteem Evolution 76
4.1.2.1 Discussion 76
4.1.3 Taking Reliability into Account 81
4.1.4 Simulations for Evaluating the use of a Reliability Stamp . . 81
4.2 ANALYZING THE CENTRALITY OF THE NODES 85

4.2.1 Eigenvector Centrality . . . 86
4.2.2 Simulations for Estimating the Probability of Prevalence . . . 86
4.2.2.1 Computational Simulations . . . 87
4.2.2.2 Laboratory Experiments . . . 90
5 CONTROLLING THE SPREAD . . . 95
5.1 BACKGROUNDS ON CLUSTERING . . . 97
5.1.1 K-means Clustering . . . 97
5.1.2 Fuzzy c-means . . . 98
5.1.3 Spectral Clustering . . . 99
5.2 COMMUNITY IDENTIFICATION ALGORITHM . . . 100
5.3 SIMULATIONS FOR SPREAD CONTROL IN A NETWORK . . . 101
5.3.1 An Illustrative Example . . . 101
5.3.2 Comparative Analyses . . . 105
6 CONCLUSIONS . . . 112
6.1 CONTRIBUTIONS . . . 114
6.2 LIMITATIONS . . . 115
6.3 FUTURE WORKS . . . 116
6.4 PUBLICATIONS . . . 117

References . . . 119

1 INTRODUCTION

In this first chapter, we present the global scenario and the context in the scientific community that motivated us to study the problem addressed in this thesis. Also, we introduce our proposal and how it differs from the approaches currently being used in the specialized literature. Further, we explain briefly the methodology employed and show the objectives defined for this research work. Finally, we give an overview of the structure for the following chapters.

1.1 PROBLEM, CONTEXT, AND MOTIVATION

The present research study was initially motivated by the problem of misinformation spreading in Online Social Networks (OSNs). There is a growing tendency for publications in this area. A search in Scopus by the term “misinformation” shows 11389 results in the past two decades, such that 75.66% of them were published from 2018 to 2022. This period of increasing interest followed the tendency of using misinformation in social media for political ends that started to cause a stir since the UK vote to leave European Union (Brexit), soon followed by 2016 USA presidential election (KUCHARSKI, 2016; BOVET; MAKSE, 2019). Filtering results by the terms “election*” or “politic*” reveals that 48.58% of them were related to these terms in the last 5 years.

The amount of publications associated with “misinformation” has increased rapidly in the last few years, roughly doubling from one year to the next. By the beginning of 2023, we see a little decrease in the slope of this ascending interest. However, the number of works in 2022 were still greater than the number in 2021. This shows that the interest in the subject is still growing.

The spreading of misinformation related to the CoronaVirus Disease 2019 (Covid-19) is one of the factors that increased the interest in the last three years. Filtering results by the terms “covid” or “coronavirus” shows that 58.4% are addressing these terms in the period from 2020 to 2022. The number of publications per year found by searching the term “misinformation” in Scopus, without filters, is depicted in Figure 1. In Figure 2, we present the results from applying the filters mentioned previously.

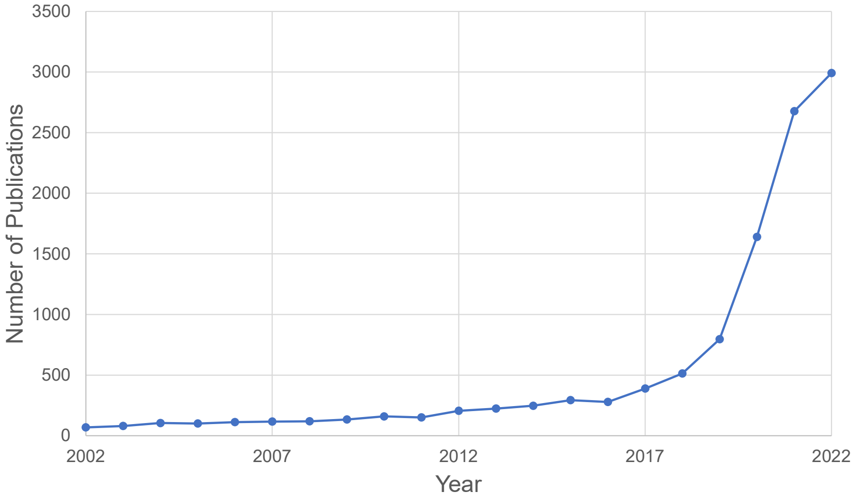


Figure 1 – Number of publications per year associated with the term “misinformation” in the Scopus platform.

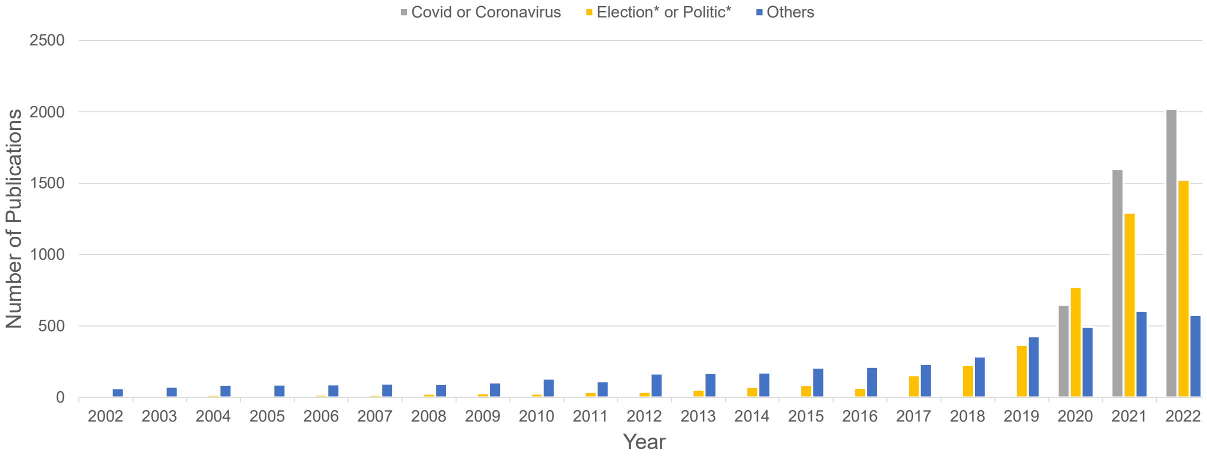


Figure 2 – Number of publications per year that relate misinformation with politics, Covid, and others.

This problem is also a concern to the press in general since false news distort the way people see the world (WATSON, 2016), and influences in crucial decision makings, such as those related to politics and health (LEWANDOWSKY; ECKER; COOK, 2017). There are initiatives for fact-checking and content creation aiming to refute misinformation, but these contents do not spread so easily as false news (VOSOUGHI; ROY; ARAL, 2018).

The Systems & Control community also shows concern about this matter. In 2017, during the International Federation of Automatic Control (IFAC) World Congress, a panel discussion occurred addressing the control challenges for social systems. At the time, Perruquetti (2017), the president of IFAC Technical Committee on the social impact of automation (TC 9.2), presented that one of these challenges is handling humans as part of control systems, and particularly the use (or abuse) of social networks to control public opinion.

The work from Lamnabhi–Lagarrigue et al. (2017) thoroughly rises great challenges for the Systems & Control community in the future. The social and techno-social networks field is pointed out as one of high research potential. Also, Lamnabhi–Lagarrigue et al. (2017) state that the success of Systems & Control theoretic methods in the analysis and control of social and techno-social networked systems requires further convergence between Systems & Control, engineering, and social sciences, to facilitate the development of new dynamic models and data analysis algorithms for testing them.

During the 2020 Virtual IFAC World Congress, Dabbene et al. (2020) presented a tutorial session entitled “Dynamical Networks of Social Influence: Modern Trends and Perspectives”. They outlined three directions for research: dynamics of opinion formation in influence networks; dynamics of interpersonal appraisals; and algorithms of identification and learning for analyzing the network’s structure.

The subject is still a trend in the community, which can be seen on some of the open invited tracks of the 2023 IFAC World Congress held this year in Yokohama. For instance, the track on “Control for Socio-Technical Network Systems” (IFAC 2023-YOKOHAMA, 2020), which is interested in contributions addressing the effects of online social media platforms over the processes of opinion formation and epidemics spreading, and also their role in inducing extreme behaviors.

During the mentioned Congress, the IEEE Control Systems Society (CSS) presented a document entitled “Control for Societal-Scale Challenges: Road Map 2030”. We highlight two points from this document that directly relates with our work. Work, Wit, and Sinopoli (2023) state that networked dynamical models are used to understand collective behaviors emerging from the interactions between single individuals, highlighting opinion dynamics and collective decision-making over social networks as one notable example where these models can be applied.

In the same road map, the article by Fay, Hosoi, and Jones (2023) sheds light on the problem of modeling and controlling pandemics. The authors state that beyond understanding the dynamics and designing control laws, it is necessary to know how physical processes depend on human behavior, and how spreading information affect this behavior. They also highlight the need to understand network effects on these dynamics, as network structures play a fundamental role in the spreading of diseases.

Information spreading is analogously associated in the literature to virus epidemics in susceptible populations (GROSS; D’LIMA; BLASIUS, 2006). However, these models simplify people’s behavior in the network and how they interact with each other to a few parameters defining the “virality” of a given piece of information.

During the development of this work, a worldwide threat started: the Covid-19 pandemic. The scientific community turned its efforts to provide predictions on the virus spreading, which could help in decisions related to the measures aiming to suppress new infections.

The models most widely used to predict epidemics' evolution date back to the work from Kermack and McKendrick (1927). They belong to the class referred to as compartmental models. In these, the population is divided into compartments, and assumptions are made about the nature and time rate of transfer between compartments (BRAUER, 2008).

These models are described by differential equations, enabling ease of utilization for various applications. On the other hand, they are lumped-parameters representations of a complex and distributed system. This means that many simplifications are made to construct such models, which could imply crucial information being left aside and limit the scenarios where the models are accurate. As an example of these limitations, the seminal compartmental models can not represent a second wave of infections during an epidemic. Roberts et al. (2015) states that many infectious diseases are fundamentally stochastic processes based on the individuals, i.e., they are more naturally represented by stochastic models, and approximating them by deterministic approaches is challenging.

In the Covid-19 pandemic, compartmental models were broadly used to predict the spreading of the disease (MEDREK; PASTUSZAK, 2020; MANDAL et al., 2020; SCALA et al., 2020). Joseph T Wu, Kathy Leung, and Gabriel M Leung (2020) explored metapopulation compartmental models, in which the population is divided into some discrete patches representing spatially distant populations or classification by age, gender, or any other important factor.

Estrada (2020) list the main approaches to forecast the progression of an epidemic at a global and local scale, being them: deterministic, data-driven (HUANG, R.; LIU, M.; DING, 2020), stochastic (HE, S.; TANG; RONG, 2020), agent-based (CUEVAS, 2020), and their combination.

A more microscopic model could be useful in generating better predictions related to the virus spreading, since it would be possible to add important factors in the analysis, such as isolation level, hygiene habits, number of hospital beds needed, among others (MANZO, 2020). We take that opportunity for using our simulation model, based on social interactions between individuals, and the knowledge acquired until that point to contribute also in the epidemiology field.

The clustering of an Online Social Network (OSN) seems to be crucial for the efficiency of recommendation algorithms in selecting content that will hold the user's attention for longer periods of time. Also, identifying the communities within a social network could be a key asset for controlling the spreading, advantaging on that knowledge to maximize the reach of a given information.

1.2 PROPOSAL AND NOVELTY

The literature suggests that simple models, based on compartmental modeling, are able to generate meaningful predictions to the spreading of a virus in susceptible populations. Furthermore, these models should be useful to also represent the spreading of information in networks, such as false news being shared through social media and instant messaging applications.

On the other hand, recent studies indicate the influence of network structure in these spreading processes. Even though some models, like the one presented by Van Mieghem (2011), started to take the network topology into account, they still simplify the agents' role in the process. In these models, it is not possible to define individual behaviors and evaluate the impact they have on the spreading process.

Therefore, this research work proposes the development of a simulation model for the spreading process taking place in complex networks, such that the agents enrolled possess behaviors that may evolve differently over time for each agent.

Humans perform a complex decision process when it comes to accepting (and possibly sharing) or not a received information, which is impacted by many factors. For instance, personality traits and self-esteem are individual characteristics that may change over time and plays a relevant role in the information diffusion in social networks.

To list a few individual behaviors relevant to the epidemiological analysis, we have the use of masks, social distancing observation, number of daily interactions, and current disease stage. These behaviors impact how much each person is able to effectively spread a virus.

Hence, our proposed model has a microscopic point of view, focusing on the behavior of the agents but also on the structure of the network connecting them. The aim of this approach is to bring more detailed information and to consider more characteristics observed in the real world.

The models used in the opinion dynamics literature already depend highly on the network structure. However, these models are grounded on the assumption of real-valued opinions, such that one agent updates its opinion by computing a weighted average of the opinions to that it has access. This type of model is not useful for the problem we are addressing, since our goal is to assess the spreading of a given piece of information.

Also, we propose a clustering technique to identify groups (communities) in social networks. We use this knowledge on the network structure to determine which are the nodes that will spread a piece of information more easily. Finding these agents can guide a synthesis process for the topology, aiming to counteract the spreading of false (or malicious) information or to enhance awareness about a real and urgent one.

1.3 OBJECTIVES

The main objective of this thesis is to develop a simulation model for social interactions occurring in complex networks of agents with evolving behaviors.

Based on that goal, this research has two major contributions: improving the quality and extent of analyses on social-dependent spreading processes, and providing policies for the synthesis of a network structure that is able to control these processes.

1.3.1 Specific Objectives

To achieve the general objective, we set the following specific objectives:

- Develop a model for the social interactions between interconnected agents;
- Modify an information exchange model for handling finite sets of information;
- Propose network structures that represent the scenarios of interest, founded on models of complex networks from the literature;
- Identify relevant features that impact the information spreading beside the network structure;
- Build models that allow dynamic matrices of interaction probability to incorporate the identified features;
- Construct a model for the spreading of communicable diseases based on social interactions;
- Create a stochastic model for the evolution of diseases in each individual that can use real-world data to adjust its parameters;
- Simulate all the proposed models in meaningful scenarios to assess their capability of improving the analyses possibilities;
- Study characteristics from the network structure that may influence the spreading processes;
- Adapt a clustering algorithm to identify communities in social networks;
- Advantage on the knowledge acquired from these characteristics to propose policies for controlling the spreading.

1.4 METHODOLOGY

The main method used to test our models, algorithms, and policies is computational simulations. The common point from the applications in information diffusion and epidemiology is that both are complex systems, influenced by the agents' behaviors and by how they interact. The research can bring a generalized model that is employed in both applications and others with similar features.

The analysis of interconnected complex systems usually requires some simplifications to be feasible (STROGATZ, 2001). To evaluate the effect of impacts related to the network topology, the agents' dynamics are relaxed (e.g., nodes with single integrator dynamics). On the other hand, when analyzing the effect of each node's behavior, the network is simplified (e.g., considering a full-connected network). We suggest a model that combines both a complex structure for the network and agents that evolve their behavior over time.

Of course, many aspects will still be neglected to permit the carrying of simulations. However, we believe that combining the complexity of these two points in a model can bring benefits to analyzing such systems. Also, its complexity is parsimonious, considering the current computational power available.

A representation of social networks by dynamic models of information diffusion combined with the gossip-based consensus algorithms is proposed (SALEM; TCHILIAN; MORENO, 2019). The community working on Opinion Dynamics is also concerned with social network analysis. However, their models assume the opinion as a real number (PROSKURNIKOV; TEMPO, 2017; JIA et al., 2015) and this would not be useful, since we aim to evaluate how a given piece of information is shared through the network.

Therefore, we provide a model for social interactions that handles discrete information, i.e., the information should not be altered during the process, it must only be exchanged between the agents. Also, it can be disassembled into two parts: the mechanism determining when an interaction occurs; and the one describing how the agents will behave whenever they are enrolled in an interaction.

The use of *social media bots* is evaluated together with their influence on two crucial aspects: the self-appraisal of nodes representing human users and the consensus of network nodes on an issue. The goal is to verify to what extent some policies could be implemented in order to tackle ethical issues in OSNs.

The problem of adding the information content's influence on the spreading process is also addressed. On that matter, we propose the implementation of a reliability stamp for each information as a policy to counteract the spreading of false/unverified content. This directive shows effectiveness to a certain point, such that it is crucial to also implement actions to detect and exclude bots from the networks.

The prevalence of a given information in the network is predicted according to the existing links between the agents. This prediction is assessed both by computational and laboratory experiments with simple robots. From these experiments, we verify that the eigenvector centrality metric carries relevant information for predicting the non-deterministic behavior of the information exchange model. Our research group has a background in cooperative robotics. Therefore, we also envision applications of the discrete opinions' model to robotic teams.

Based on spectral clustering and fuzzy c-means, we build an algorithm for clustering the nodes of a network. This procedure is known as community identification in the specialized literature. We propose policies for choosing possible spreaders that would make an information disseminate the fastest. To verify the efficiency of these policies, we perform comparative tests between them and another one chosen as the benchmark.

Social interaction is the main source of infections in pandemics such as the one caused by the Severe Acute Respiratory Syndrome CoronaVirus-2 (SARS-CoV-2). Hence, with the onset of Covid-19 pandemic, we were aware of the connections and similarities of our analysis subject and the problem of predicting a virus/disease spreading. For this purpose, we formulate a model for the agents that describes them as a set of states with probabilistic transitions. The model is flexible enough to adjust for other sets of states and transitions.

So, the same simple model for social interactions from the other case study is used, allowing space to use complex networks for connecting individuals, mainly based on the Watts-Strogatz model. Flexibility in the network structure of social interactions allows the analysis of more refined Non-Pharmaceutical Interventions (NPIs).

Also, a generic microscopic (i.e., at an individual level) infection evolution model is proposed. We believe that this is the first time that a model considers non-deterministic transitions between stages of a disease, in such a way that the probability of occurring a transition is time-dependent on the last transition. Further, we provide the definition of the transition probabilities for the scenario addressing the resumption of on-site teaching activities. These probabilities are established considering a disease modeled with 4 stages (Susceptible-Exposed-Infected-Removed), and a fast screening and isolation of infected individuals.

The individuals have their behavior defined according to their current state, and they transition between states by two distinct processes: social interaction (for transitioning from a Susceptible state) and infection evolution (for transitioning from any state after the infection). The model allows the definition of individual and time-varying behaviors for the agents and describes the pattern of social interactions between them.

1.5 CONTRIBUTIONS OF THE THESIS

In this section, we present a summary of all the contributions of envisioned in this thesis. To help the reader keep track of the subjects addressed in this work, we illustrate how every contribution are linked to each other in Figure 3.

The common point between all contributions is the model of social interactions over networks of interpersonal relationships. In the second level, we divide the models by the interactions' consequences, which are the information exchange in online social networks and the infection by a disease in a given population. The first case study

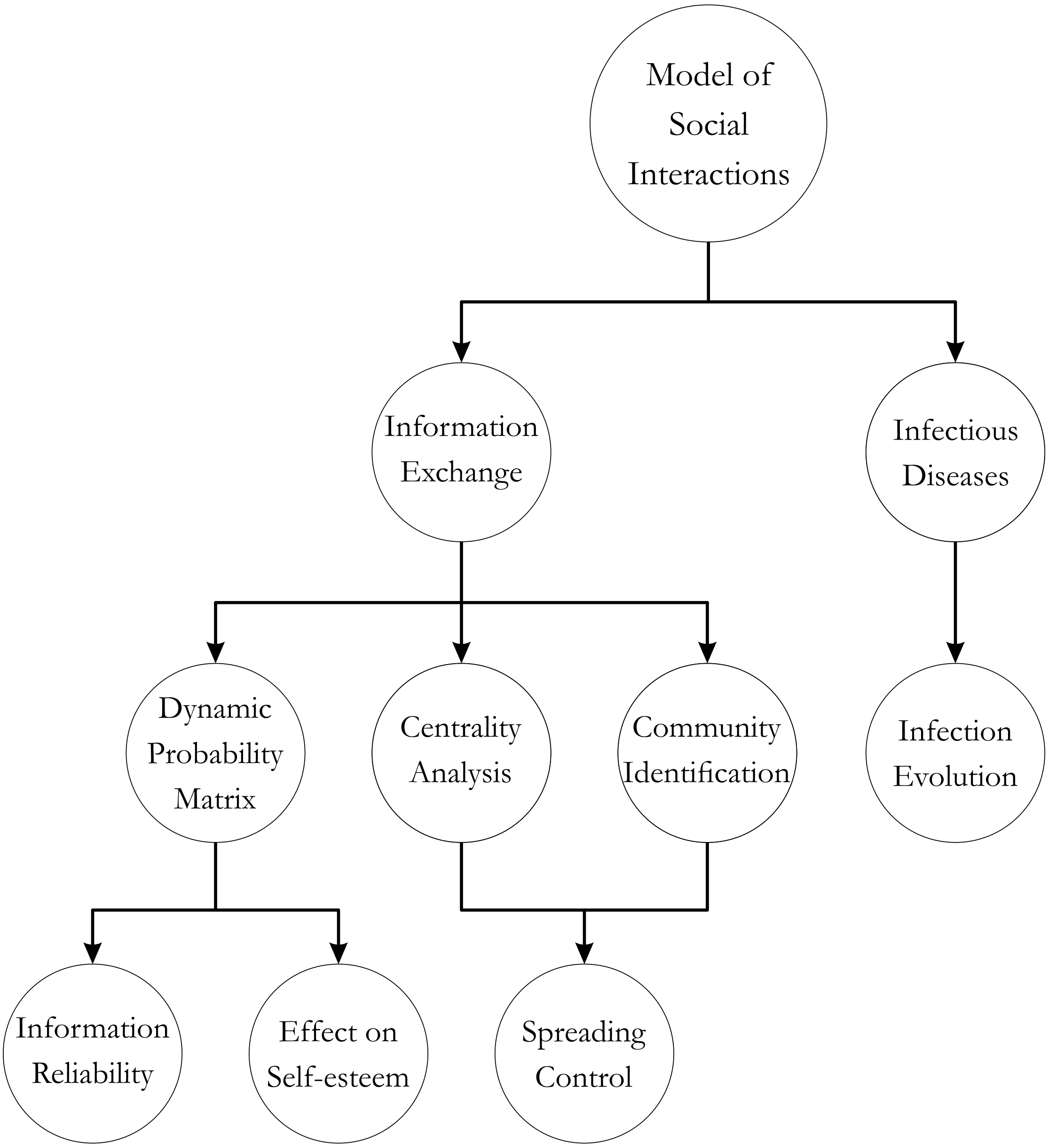


Figure 3 – Overview of the contributions of the thesis and how they are connected to each other.

allows us to contribute both on analysis and synthesis, while the second one is focused on analysis, since imposing the network structure for interactions in the physical world would incur in many ethical issues.

Following the case study over infectious disease's branch, in the next level, we propose a stochastic model for the disease's evolution in each individual. Namely, we describe how the person, after getting infected, will pass through each stage of the disease. The transitions between stages are based on time-dependent probabilities.

On the information exchange's branch, we provide models that describe how the information reliability and the individual's varying self-esteem would impact the spreading processes. These two contributions have in common the fact that they dynamically change the probabilities of happening a social interaction between two individuals.

Still on that case study, we propose a policy for choosing the nodes that will spread the information faster. The same policy can be used for identifying the spreaders and counteracting the diffusion of false or malicious information. This contribution relies on two previous ones: the use of centrality metrics as analysis tools for understanding the spreading process; and the algorithm for identifying communities in social networks.

1.6 DOCUMENT STRUCTURE

The rest of this document is organized as follows. Chapters 2 to 5 are self-contained, bringing all information needed to understand the results presented in each of them. In Chapter 2 we present our model for social interactions over a network. After, in Chapter 3, we propose an infection evolution model for the agents and use it together with the social interactions model to analyze the spread of diseases in susceptible populations. Another application for the social interactions model is the spreading of information in Online Social Networks, which is presented in Chapter 4. In this chapter, we also bring the modeling of two aspects that make the matrices of interaction probability dynamic: the humans' self-esteem and the information's reliability. To close the analytical part of the information exchange study, we examine the relation between a centrality measure and our model. Further, in Chapter 5, we bring an algorithm for community identification in social networks. We use the output of this algorithm together with previous knowledge on centrality metrics to propose policies of network synthesis for controlling the spread of information. Finally, Chapter 6 summarizes the conclusions obtained during this research work, along with its limitations and proposals for future works.

2 MODELING SOCIAL INTERACTIONS

The ability to establish complex social networks is a distinct skill from humanity; for many researchers, it is a milestone for human evolution. From small-group organization to the foundation of empires with complex lines of communication, from the diffusion of information in a tribe to global communication, the evolution of social networks could not be separated from technological transformations.

Since the beginning of the modern studies of sociology, in the nineteenth century, society was viewed as an interconnection of social actors. In the 1930s, a significant contribution was the work from Jacob Levy Moreno, whose theory of society focused on the networks of interpersonal relations that join individuals (FREEMAN, 2004). From the earliest works, ethical issues have played a central role in the development of theories on the behavior of society, whether in the assumption of the behaviors of specific social groups or in proposing tools to control and influence public opinion (BERNAYS, 2005; LIPPMANN, 2017).

In addition to the evolution of studies in social networks, the development of new mass media has impacted the way individuals interact socially, increasing the reach of the media to a global scale. These new *media* also gave rise to new tools to interfere and influence society. In addition to the increase in the range and speed of information dissemination, new technologies, such as press, radio, television, have defined new social relationships. According to influential studies developed in the 1960s, the medium defines new relationships, being itself a message, as Marshall McLuhan stated: "*the medium is the message*" (MCLUHAN, 1994).

With the evolution of the media and its uses, regulations have been developed to prevent distortion and abuse, in marketing, public opinion manipulation, and respect for privacy. In the case of influence on electoral processes, various constraints and practices have been regulated around the world (EODS, 2016). The principles underlying these initiatives are the guarantee of the right of expression of the various social players involved in the processes, avoiding asymmetries, curbing the abuse of economic power, preventing the dissemination of false information, and avoiding the manipulation of public opinion.

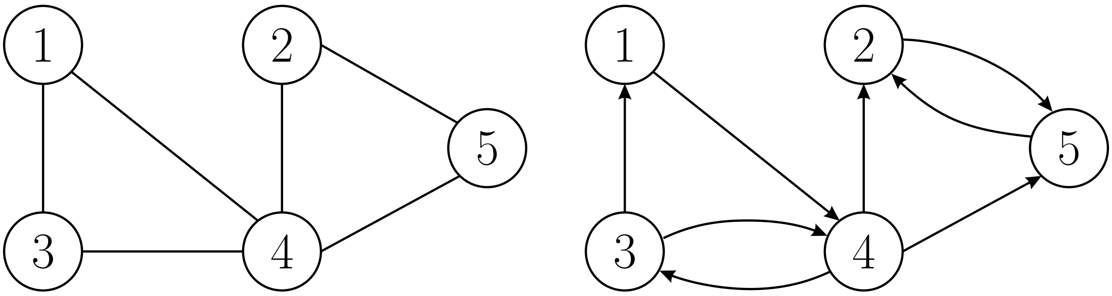
In this chapter, we propose an approach for modeling the social interactions between humans. This model dictates when an interaction occurs based on probabilities that are intrinsic to the network formed by the humans' social ties. We can use the same model for different network structures, allowing to cover both real-world and virtual interactions.

The model presented here does not describe what happens when an interaction occurs. Hence, we can use the same model for both the applications of information spreading and epidemics. This flexibility is what allows us to define the behaviors of each agent in the network on a microscopic level, accordingly to the application and scenario being studied.

2.1 MODELING NETWORKS

Graph theory is the common ground for every network representation presented in this work. Hence, a graph $\mathcal{G}$ is constituted by a set of *vertices* or *nodes*, $\mathcal{N}$, a set of *edges*, $\mathcal{E}$, such that $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$, and also an incidence function that joins every edge to two vertices. If the vertices (or ends) of an edge are different then it is called a *link*, otherwise it is called a *loop* (LUCCHESI, 1979; TUTTE, 1966).

It is also valid mentioning that graphs can be directed or not. In undirected graphs, there is no distinction between the ends of an edge. On the other hand, in directed graphs, the edges (also called *arcs* in this case) have a direction indicated by an arrow coming from the origin to the destiny node (BONDY; MURTY, 2008; GOLDBARG, M.; GOLDBARG, E., 2012). Figure 4 shows examples of both types of graphs.



(a) Undirected (b) Directed

Figure 4 – Graphs examples.

Both graphs shown in Figure 4 have five vertices. In the left, Figure 4a, a undirected graph is depicted with six edges. In the right, Figure 4b, a directed graph with eight arcs is illustrated. A directed graph can also be called *digraph*.

We can use both types of graphs to represent interactions in networks. Each agent involved is considered as a node. For undirected graphs, each edge represents a two way communication between the nodes joined by it. However, in the directed case, the arc indicates explicitly that communication occurs only in one way. The origin and destiny nodes can be called conveniently as transmitters and receptors, respectively (REN; BEARD, 2008).

Also, there are algebraic representations for graphs, such as the *adjacency matrix*. Let a undirected graph $\mathcal{G}$ formed by n vertices, its adjacency matrix is defined as $\mathcal{A} = [\alpha_{ij}], i,j = 1 \ldots n$. The elements α_{ij} indicate the existence of edges linking nodes i and j. Hence, if there is an edge connecting vertices i and j then they are called adjacent or *neighbors* (WILSON, 2010).

Let a directed graph $\mathcal{D}$ with vertices set $\mathcal{N}$, such that $|\mathcal{N}| = n$, and a set of arcs $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$. Its corresponding adjacency matrix is of the order $n \times n$ and is represented as $\mathcal{A} = [\alpha_{ij}]$. The element α_{ij} equals 1 if the arc (j,i) is in the set $\mathcal{E}$ of edges, otherwise it equals 0. That means, $\alpha_{ij} = 1$ if there is an arc in the direction from j to i in the digraph (LI; DUAN, 2014).

The adjacency matrices in (1)a and (1)b are equivalents to the graphs shown in Figures 4a and 4b, respectively. It is noticeable that adjacency matrices will always be symmetric in undirected graphs. Following this concept, we can define undirected graphs as a special case of directed graphs, in which for every arc (i,j) there will always be the arc (j,i).

$$\mathcal{A}^{(a)} = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix} \tag{1)a}$$

$$\mathcal{A}^{(b)} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix} \tag{1)b}$$

In a undirected graph, the *degree* $d(i)$ of node i is the number of edges connected to it (GODSIL; ROYLE, 2001). This value can be obtained from the adjacency matrix, such that $d(i) = \sum_{j=1}^{n} \alpha_{ij} = \sum_{j=1}^{n} \alpha_{ji}$. In the case of digraphs, there is a difference between in-degree, $d_{in}(i)$, and out-degree, $d_{out}(i)$, for each node $i \in \mathcal{N}$. The in-degree indicates how many arcs arrives at i, it can be calculated as $d_{in}(i) = \sum_{j=1}^{n} \alpha_{ij}$. On the other hand, the out-degree represents the number of arcs departures from i, being evaluates by $d_{out}(i) = \sum_{j=1}^{n} \alpha_{ji}$ (WILSON, 2010).

2.2 COMPLEX NETWORKS

Complex networks are networks with a more complex structure than classical graphs (e.g., full, lattice, ring, star) and fundamentally random graphs (e.g., Poissonian degree distribution). Many real-world networks are proved to be complex networks. This implies skewed distributions of connections, high clustering, and non-trivial temporal evolution. Also, in despite of their high number of vertices, these architectures present a small degree of separation between the nodes (DOROGOVTSEV; GOLTSEV; MENDES, 2008).

Social networks are important case studies in the theory of complex networks and multi-agent systems because they differ from many natural and human-made complex networks in the sense that they do not present a cooperative behavior. The OSNs also work as tools for mapping the wiring diagram behind our social system, which is a necessary step to describe its complex behaviors (BARABÁSI et al., 2016). The mathematical models for representing this kind of networks need to be simple enough to be examined, but able to capture the complex behavior of real social groups (PARSEGOV et al., 2017).

Watts and Strogatz (1998) state that the topology from networks of coupled dynamical systems is usually assumed to be completely regular or completely random. Hence, they introduce a model of networks that can be tuned in the middle ground between these two extremes. They name these networks as “small-world” due to the small-world phenomenon, also known as six degrees of separation.

The procedure for creating a small-world network is dependent of a single parameter. It starts from a regular ring lattice, where every node is connected to its closest neighbors only. Then, every edge can be randomly “rewired” with probability p, avoiding duplicates. Figure 5 depicts the outputs of this process as the value of p changes.

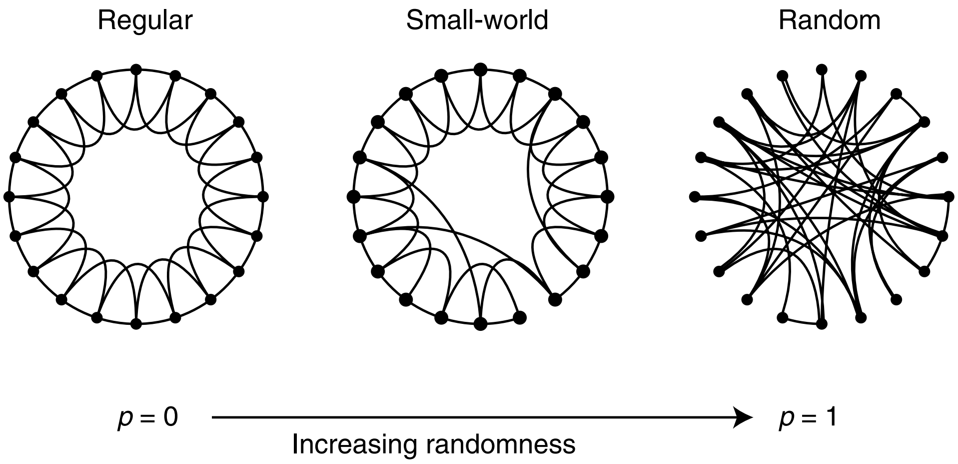


Figure 5 – Random rewiring procedure, from a regular ring lattice to a random network.
Source: Watts and Strogatz (1998)

According to Strogatz (2001), the slightest bit of rewiring transforms the network into a “small-world”, with short paths between any two nodes (like in a completely random graph) and maintaining the network highly clustered (as in a regular lattice). These few “long-distance” connections are responsible to increase the speed of spreading (of diseases, information, etc.) in this type of network (WATTS, 2004). These links are the “weak ties” that Granovetter (1983) addressed in his work.

This network topology is particularly interesting for modeling the epidemic evolution of infectious diseases, such as the Covid-19. It can bring more realistic predictions about virus’ spreading. Manzo (2020) suggests that considering complex social networks can enable more sophisticated interventions that apply to specific categories or sets of individuals, expecting collective benefits. This contribution should come because classical compartmental models look at virus diffusion without modeling the topology of social interactions.

2.3 CONSENSUS PROBLEM

The model proposed in this thesis for describing social interactions between agents is founded on a class of problems known in the literature as consensus. In this section, we present the seminal formulation of the consensus problem to help the reader understand our proposal.

Consensus problems can be represented as directed graphs, where agents are represented as nodes and their interaction as edges. Let the set of all communication nodes be denoted by $\mathcal{N}$ and the set of edges (or links) by $\mathcal{E}$, where $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$ such that $(j,i) \in \mathcal{E}$ if node i receives information from node j. The set of the in-neighbors of node i (including i itself) will be called $\mathcal{N}_i$, i.e., $\mathcal{N}_i = \{j \mid (j,i) \in \mathcal{E}\}$. (MESBAHI; EGERSTEDT, 2010; CARVALHO; CORREIA; MORENO, 2015)

The continuous-time consensus algorithm is given by (2) (REN; BEARD; ATKINS, 2007), where x_t^i is the vector of information states of node i and α_{ij} are the elements of the adjacency matrix, which were described previously.

$$\dot{x}_t^i = \sum_{j \in \mathcal{N}_i} \alpha_{ij}(x_t^j - x_t^i) \tag{2}$$

Since the communication between agents occurs at discrete instants of time, the information state will be updated by a difference equation and the discrete-time consensus algorithm has the form (REN; BEARD; ATKINS, 2007):

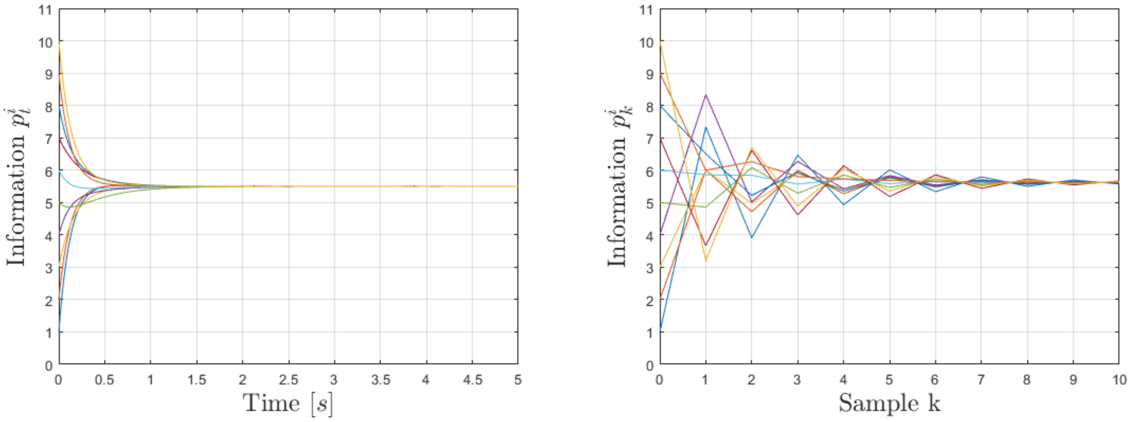
$$x_{k+1}^i = \sum_{j \in \mathcal{N}_i} \pi_{ij} x_k^j \tag{3}$$

with $\pi_{ij} \geq 0$ and

$$\sum_{j \in \mathcal{N}_i} \pi_{ij} = 1, \forall i \in \mathcal{N} \tag{4}$$

where the terms π_{ij} are called consensus weights.

To illustrate, we apply both continuous and discrete-time versions of the algorithm for randomly defined networks of 10 agents. The results are depicted in Figure 6, where each line represents the information of one agent evolving in time. The agents' initial information is their own index, i.e., $x_0^i = i$. As we can see, there is convergence on both versions of the algorithm, since the network topology enables it.



(a) Continuous-time (b) Discrete-time

Figure 6 – Information of each node evolving in time according to the consensus model.

2.3.1 Opinion Dynamics

The model presented by DeGroot (1974) describes the evolution of opinions in a group of individuals trying to reach consensus on a given topic. It assumes that individuals update their opinions as convex combinations of their own and those displayed by others (JIA et al., 2015). So, the opinions update as

$$z_{k+1} = \mathcal{W} z_k \tag{5}$$

where $z_k \in \mathbb{R}^n$ is a vector holding the opinions of the n agents; $\mathcal{W} = [w_{ij}]$ is a row-stochastic matrix that describes the influence weights.

So, each edge, $j \xrightarrow{w_{ij}} i$, represents the weight that individual i accords to the opinion from individual j or how much influence j has over i. Notice that, mathematically, this model is identical to the discrete-time consensus algorithm. The difference comes from the meaning attached to the variables.

Proskurnikov and Tempo (2017, 2018) have done incredible work by highlighting the most classical models of social dynamics and also bringing the more recent models developed. Much of the theoretical knowledge needed to understand these models are compiled by Bullo (2019) in his book.

2.3.2 Gossip-based Consensus

The asynchronous gossip-based consensus algorithm (BOYD, S. et al., 2006) is widely explored in the literature. In which, a pair-wise communication is considered, such that, at each iteration k, an active agent i_k is randomly selected following a uniform distribution.

Then, the active agent has probability p_{ij} of interacting with agent j. These probabilities of interaction between agents are arranged in an $n \times n$ row-stochastic matrix,

$$\mathcal{P} = [p_{ij}]$$

where n is the number of agents in the network.

The gossip algorithm also defines how the agents update their information, but this will be discussed comprehensively in the Section 4.1. Here, the main interest is the mechanism that selects when an interaction between two agents occur.

In the synchronous version of this algorithm, no active agent is selected. Instead, at each step k, all the agents update their information following the probability distribution imposed by the matrix $\mathcal{P}$.

2.4 MODEL FOR SOCIAL INTERACTIONS

Finally, we present our proposed model for the manner people interacts with each other, based on ideas coming from the two models shown in the previous section, which are widely used in the literature on consensus and opinion dynamics.

They will be called, for short, gossip and DeGroot models, and we present how the relation between them can be established. This link is made by defining the matrix of probabilities of interaction equal to the matrix of influence weights, i.e.,

$$\mathcal{P} \triangleq \mathcal{W} \tag{6}$$

That connection between the models is the same as saying that the probability of agent i to interact with agent j is proportional to the influence that j has over i. The matrix of probabilities now represents the trust of the agents in one another.

It is noticeable that this modeling for the social interactions is highly dependent of the network topology. Therefore, a good representation for the connections between the nodes in an specific application is crucial to achieve meaningful results. Although there are models in the literature for representing complex networks, the identification of structures in social media is a very active area in many fields of research (PROSKURNIKOV; RAVAZZI; DABBENE, 2018).

Proskurnikov, Ravazzi, and Dabbene (2018) points out that, in order to model real interaction in social media, it is necessary to account the time-varying nature of such networks. Epidemics modeling, the other application addressed in this work, could also benefits from the use of a time-varying structure in the analysis of specific scenarios. This is the case for some of the simulations that are presented in the upcoming sections.

3 DISEASES SPREADING

Mathematical models are used to predict epidemics' evolution since the beginning of the 20th century, when the seminal work by Kermack and McKendrick (1927) gave the first step to the class known as compartmental models. They are based on differential equations that describes the nature of how the population is transferred between compartments.

Compartmental models were broadly used to predict the spreading of the Covid-19 (MEDREK; PASTUSZAK, 2020; MANDAL et al., 2020; SCALA et al., 2020). However, Roberts et al. (2015) states that many infectious diseases are fundamentally stochastic processes based on the individuals. The work by Estrada (2020) brings a compilation of the main approaches on the forecast of epidemics, being them: deterministic, data-driven (HUANG, R.; LIU, M.; DING, 2020), stochastic (HE, S.; TANG; RONG, 2020), agent-based (CUEVAS, 2020), and their combination. Further, some interesting studies (LIU, Y., 2020; MARFAK et al., 2020) used Markov chains to model and predict the spreding. Also, diverse ways of social contact can cause the transmission of a disease, such as social contact networks, transport systems, metapopulations, etc.

Hunter, Mac Namee, Kelleher, et al. (2020) present a hybrid model, combining agent-based modeling with classical compartmental models to reduce the computational cost of the simulation. Their main assumption is that analyzing at a microscopic level is more relevant when only a few individuals are infected and, after a certain threshold is passed, individual movements should not matter much.

The urgency of the Covid-19 pandemic certainly concerned the worldwide scientific community. A search in some of the most prestigious journals can emphasize these concerns. For instance, searching for papers on epidemiological models for Covid-19 at Nature will lead to a total of 241 works, from which we highlight those by Xi He et al. (2020) and Scala et al. (2020). For more details on the disease, a search in the British Medical Journal will result in more than 3,000 results. However, we emphasize those by Wise (2020) and Mahase (2021).

Recent works present forecasting methods built over the data collected from several countries. Following, we highlight some of these, which use data from the Covid-19 pandemics in Brazil and provide results concerned with predicting the evolution in the country. Medeiros et al. (2021) propose a method based on regression models to forecast the evolution of the pandemic in countries using past data from other countries that have already passed through a similar stage. The work by Ribeiro et al. (2020) brings several approaches from Machine Learning and also presents a comparison between them in terms of accuracy. Scabini et al. (2021) provide a model that builds on complex networks with multiple layers, such that each layer represents a type of social group to which the individuals belong (family, work, religion, etc.).

Ekinci (2021) models new cases of Covid-19 by conditional variance and asymmetric effects employing variations of Generalized Autoregressive Conditional Heteroscedasticity (GARCH). A comparison between artificial intelligence models is also presented by Silva et al. (2020), being coupled with pre-processing Variational Mode Decomposition (VMD).

Drews et al. (2022) investigate the difference of forecasts produced by a compartmental model and a statistical time series model. They analyze the sensitivity of the model parameters in data from ten countries. A computational tool is presented by Paiva et al. (2021) to study the pandemic and perform a trend analysis, in which numerical optimization is used to calibrate the model parameters.

The automation and control community in Brazil is also concerned with the pandemic situation, counting with contributions of many researchers in the modeling and control of epidemics. Santos, Rodrigues, and Oliveira (2022) present an equation-based model within a sliding modes control strategy. The work of Batista and Cunha Jr (2020) focuses on regression using the available data.

Batistela, Cabrera, et al. (2020) also present an equation-based model, adding an immune compartment to enable vaccination effects analysis. The methodology brought by Tosin, Cunha Jr, and Coelho (2020) is based on global sensitivity analysis, and they use it to compare variations of a six-compartments model. Almeida, Morgado, and Arenas (2020) apply a Markov chain-based stochastic compartmental model to regionally predict the pandemic's evolution.

The computational model proposed by Gomes and Serra (2021) aims in analyzing the spread of Covid-19 by using type-2 fuzzy systems and Kalman filtering over the experimental data. Oro et al. (2020) use Autoregressive Distributed Lag Model to forecast the number of new daily cases of the disease in a specific population.

Vasconcelos et al. (2022) analyze the epidemic situation through a computational tool that fits logistic curves of growth to the observed data. The paper by Costa Jr, Martinez, and Geromel (2022) presents a probabilistic dynamic model of the SIR (Susceptible-Infected-Recovered) class that describes the temporal behavior of epidemics in discrete time.

The model presented by Dias, Queiroz, and Araujo (2022b) combines deterministic and stochastic features and allows the population stratification into subgroups. Pazos and Felicioni (2020) apply proportional control to Non-Pharmaceutical Interventions (NPIs) aiming at the prevention of Covid-19 spreading over a six-compartments model.

Purposing to avoid the health system overload, Dias, Queiroz, and Araujo (2022a) implements a proportional-integral controller on a compartmental model. Pataro et al. (2022) suggest using model predictive control to plan social distancing policies to mitigate the epidemic's spread.

Despite the variety of modeling views presented by the community, there is still a gap concerning microscopic approaches. According to Pastor-Satorras et al. (2015), we are currently in a second golden age in epidemic modeling, led by the improvement in the real-world accuracy of the models. This is being achieved by the combination of available data in large amounts and the explicit simulation of entire populations down to a microscopic scale. According to Manzo (2020), a microscopic model can add crucial factors in the analysis, such as isolation level, hygiene habits, and number of hospital beds needed.

A more recent search on the literature led us to papers that reinforce the significance and contributions of our work because they are concerned with nuances of the problem that are addressed in our proposal. Parino et al. (2021) introduce a metapopulation model to evaluate more sophisticated NPIs. Ramos et al. (2021) modify a compartmental model for including control measures, such as social distancing and contact tracing. A partial differential equations model is presented by Viguerie et al. (2021) for capturing dynamics based on human habits.

Gumel et al. (2021) present a brief overview of models used to study Covid-19, which includes agent-based, network, and statistical models. They also illustrate how to incorporate new features to seminal epidemic models, such as vaccination and NPIs. In their work, Sturniolo et al. (2021) state that effects of testing, contact tracing, and isolation at individual levels can be approximated to aggregate measurements on the population level.

The work from Nielsen, Simonsen, and Sneppen (2021) warns about the evidence mounting that the Covid-19 pandemic is characterized by “superspreading”. They also state that to capture this phenomenon an individual-scale model is needed. Mo et al. (2021) consider passenger contact in public transit to be a crucial factor in the spreading of infectious diseases. Hence, they introduce a model for this network of contacts that is time-varying. They also refer to the need of identifying “influential passengers”, which is the same as the concern with superspreaders raised by Nielsen, Simonsen, and Sneppen (2021).

After an extensive literature review, we could identify many different approaches to the forecasting of epidemic spreading. However, the complex networks governing social interactions are yet to be investigated in this context. Hence, our contribution to this field is achieved by providing a simulation model that can handle both the complex structures of social interactions’ networks and the behaviors of the agents in a microscopic level.

Our model for infection evolution eases the customization for other diseases based on available data about them. This process will be more intuitive than in other models since it is a matter of defining the relevant states of the disease and building a histogram of people transitioning between the states on each given day. This ease of adding and removing states for the infection evolution also makes more complete analyses possible. For instance, a state representing people in need of a scarce resource, such as mechanical ventilators, could enable adjusting to the forecasted demand. Other examples would be the possibility to set different behaviors for people in isolation or after getting vaccinated to see how it affects the overall spreading.

In this Chapter, we also present two applications of the simulation models. The first one is a simulation of the SARS-CoV-2 spreading in a building at UFSC (Section 3.3), and it aided in assessing the impact of resuming presential activities at the University. The second application intended to simulate the epidemic's evolution in a city (Section 3.4), generating meaningful information for the health administration entities. These results show the model flexibility to changes both in the network and in the infection evolution descriptions.

3.1 BACKGROUNDS ON MODELING EPIDEMICS

In the literature on epidemiology, mathematical models are used to simulate and analyze the evolution of epidemics in populations of interacting individuals. There are many models for a variety of scenarios and disease's types, such as SI, SIS, SIR, and SEIR. They are named according to the first letter in following terms: **S**usceptible, **I**nfected, **E**xposed, and **R**ecovered (or **R**emoved). These terms represent the individual's states considered in each model or the compartments in which individual can be placed, in a compartmental modeling point of view. The **S**usceptible compartment holds the individuals that are healthy and may be infected by the etiological agent. The **I**nfected contains individuals that currently have the disease and are transmitting it. The **E**xposed includes those individuals that were exposed to the etiological agent, but did not showed symptoms and/or are not able to transmit it yet. Finally, the **R**ecovered contains individuals that have been removed from the transmission chain of the modeled infectious disease, either by recovering and becoming immune or by death (DALEY; GANI, 1999). We believe that the term **R**emoved is more comprehensive than **R**ecovered, and it does not affect the models' nomenclature. Also, the currently available data on Covid-19 indicates that asymptomatic individuals are able to infect others, which disagrees with these classical models.

In compartmental epidemiological models, as the etiological agent spreads in the population, there is a transit of individuals from one state to another. In other words, the population's fraction allocated in each compartment changes as individuals are exposed, infected, healed or killed. The key distinctions among models are the states considered and how the transitions between them occur for each studied epidemic. The SIS (Susceptible-Infected-Susceptible) model represents diseases for which there is no acquired permanent immunity; after a person being infected and recovered, she/he becomes susceptible to the disease once more. In SIR (Susceptible-Infected-Recovered) model, once an individual is infected by the disease, she/he can only be removed from the chain of transmission. In SEIR models, there is an intermediate stage representing the latent period between exposure to the etiological agent and symptoms onset, allowing the depiction of different degrees of virality that occur in these two stages (HETHCOTE, 2009).

We show the transitions between each state in a SEIR epidemiological model in Figure 7. The Greek letters over the arrows represent the rate that individuals go (or "flow") from one compartment to the other: β is the transmission rate, with which the susceptible individuals are exposed to infected individuals; ε is the rate of exposed individuals becoming infected, the inverse of this value $(1/\varepsilon)$ is the incubation period; γ is the recovery rate, describing how "fast" an infected individual recovers from the disease; δ is the mortality rate of the disease. The inverse of the sum of mortality and recovery rates $(1/(\delta+\gamma))$ indicates the mean time of infection or prevalence of the etiological agent in the affected population.

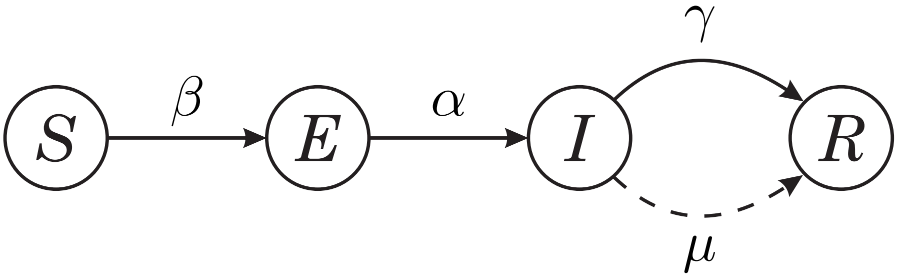


Figure 7 – SEIR model diagram.

It is common to express the SEIR compartmental model as a set of differential equations describing the evolution in time of the number of individuals in each state or compartment. For the model shown in Figure 7, we have

$$\begin{aligned}
\dot{S}_t &= -\varsigma\beta\frac{I_tS_t}{n}\\
\dot{E}_t &= \varsigma\beta\frac{I_tS_t}{n} - \varepsilon E_t\\
\dot{I}_t &= \varepsilon E_t - (\gamma+\delta)I_t\\
\dot{R}_t &= (\gamma+\delta)I_t
\end{aligned}$$

where S_t, E_t, I_t, R_t represent, respectively, the number of susceptible, exposed, infected, and removed individuals at time t; ς is the average number of contacts for each individual in the population. The total number of individuals in the population is denoted as n. Notice that $S_t + E_t + I_t + R_t = n$ for any $t \geq 0$.

Compartmental models assume homogeneous populations, where the interactions between individuals occur in a regular way. Although they can provide good estimates for the spreading of epidemics in urban and uniform populations, such approaches may wrongly estimate the evolution of infectious diseases in populations from diverse cities/countries. The scientific community is aware of these flaws and has been exploring different approaches, such as subpopulation, spatial and intertwined models.

Recent works, like the one presented by Shi, Cao, and Feng (2020), extended the SEIR model in order to encompass quarantine states and infections during the **E**xposed state. This revealed to be needed, since new infectious diseases, such as Covid-19, are presenting contagious behavior even in asymptomatic individuals. In classical models, these individuals are considered exposed to the virus, but not able to transmit it to others (LIU, T. et al., 2020).

The model proposed by Shi, Cao, and Feng (2020) to depict the Covid-19 epidemic has three new states besides the originals (SEIR). They are **Sq**, **Eq**, and **H** representing, respectively, the number of susceptible individuals in quarantine, exposed individuals in quarantine, and hospitalized individuals. The authors considered that hospitalized individuals are unable to infect others, or have a low probability of doing it, because they are isolated. The states and transitions considered for this model are depicted in Figure 8. The transit between the SEIR compartments is given by the rates β, δ, and ε described before. For the new compartments, specific isolation rates are defined to represent the transition into them.

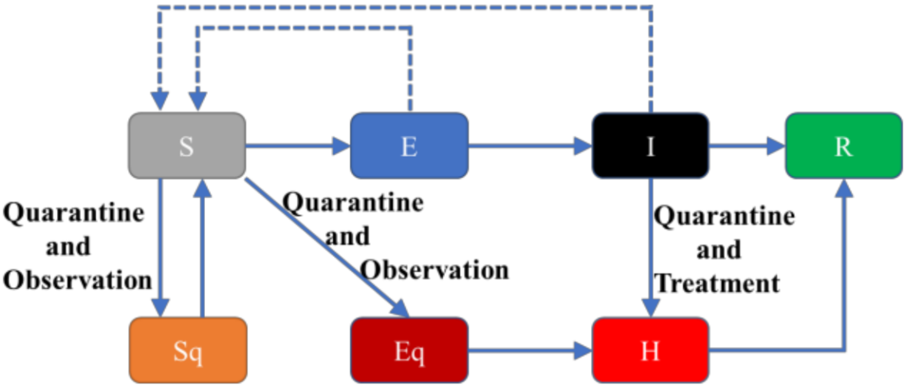


Figure 8 – Diagram for the model altered to address a Covid-19 epidemic.
Source: Shi, Cao, and Feng (2020)

3.2 SIMULATION MODEL

Each agent in the simulation is modeled as a finite set of states and the transitions between them can occur in two main ways. The first one is given by a *social interaction* process that can cause a susceptible agent to become infected. Once infected, a second process starts, which we called *infection evolution*. In this second stage, the states and transitions are defined according to the disease being represented. We depict this general view of the agent in Figure 9. We give more details on each of these processes in the following sections.

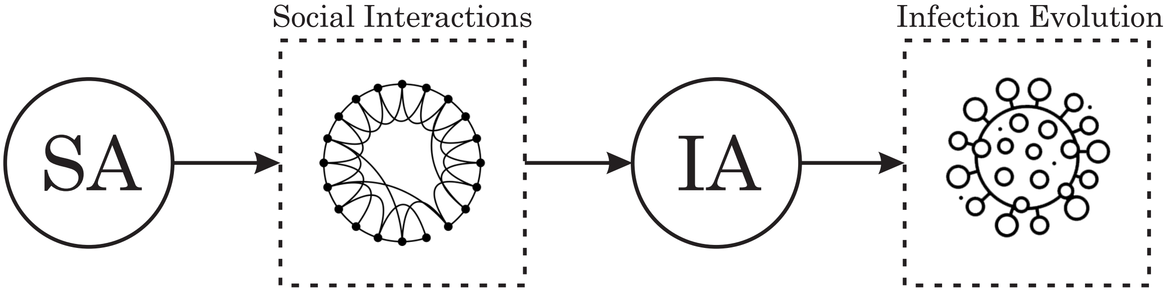


Figure 9 – Overview of the agents' evolution, where SA means "Susceptible Agent" and IA means "Infected Agent".

The social interactions can cause a susceptible agent to be infected by the disease. Besides the probability of occurring the interaction, the infection also will happen with a given probability, which can be customized according to the reproduction ratio of the studied disease. The interactions in our model are given in a pairwise and non-deterministic way, as described by the model presented in Section 2.4. The probability of two agents interacting is given by the weight attributed to the link between them in the adjacency matrix that describes the network. Hence, the configuration of the network plays a fundamental role in this model.

For this reason, we thought necessary to consider more complex network structures in our work, such as the Watts-Strogatz model presented in Section 2.2. Also, the social interaction pattern can be time-dependent to represent periods of quarantine and relaxation in the same simulation. There is room also to define diverse classes of agents, representing people that are more careful or others that have more links than average, for instance.

3.2.1 Infection Evolution Model

The infection evolution model describes the states and transitions one agent can experience after being infected. Persons react in different ways to the same diseases. Hence, we defined probabilistic transitions between states. The probability of transitioning is time-varying and depends on how many days have passed since the last transition. We illustrate that concept in Figure 10. These probabilities can be inferred from real-world data, e.g., accounting for how many people display symptoms each day after being infected by an etiologic agent.

An important remark is that the probabilities in our model are cumulative, in the sense that, for instance, even if few people transition on the 8th day, the probability of transitioning on this day is higher, if is given that the transition did not occur before. Generically, we can say that the conditional probability of transitioning on day d, given

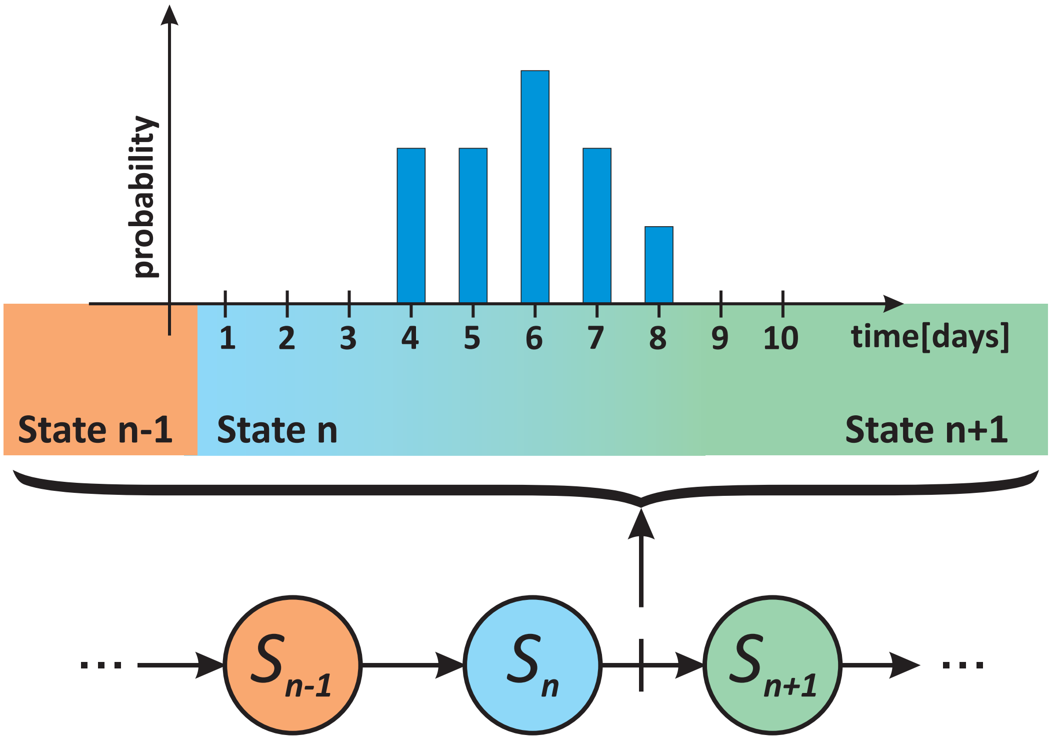


Figure 10 – Illustration for the concept of probabilistic transitions between states, in which the probability varies as time passes since the previous transition.

that it has not transitioned yet, is obtained by summing all probabilities of transitioning from day 1 until day d. This cumulative distribution guarantees that the agents do not get caught in a deadlock.

Further, if a state can lead to multiple other states, the probability functions should be multiplied by factors that add up to one. These factors represent the percentage of people that transition to each state and can also be inferred from data. For instance, in the Covid-19 pandemic, the infected agents can either present Mild Symptoms, Severe Symptoms, or Recover without showing any symptoms.

3.2.2 Simulation Steps

In this section, we summarize the generic steps of a simulation for the evolution of an epidemic using our model. We present a flowchart for these steps of our simulation model in Figure 11.

The initial step is the Scenario Set-up, consisting of configuring:

- The Infection Evolution Model;
- The Social Interaction Model.

To define the infection evolution model, we need to describe which states an agent can go through and how (or if) one state can lead to another. This definition should be reasonable following the disease of interest. Also, in this step, the probabilities

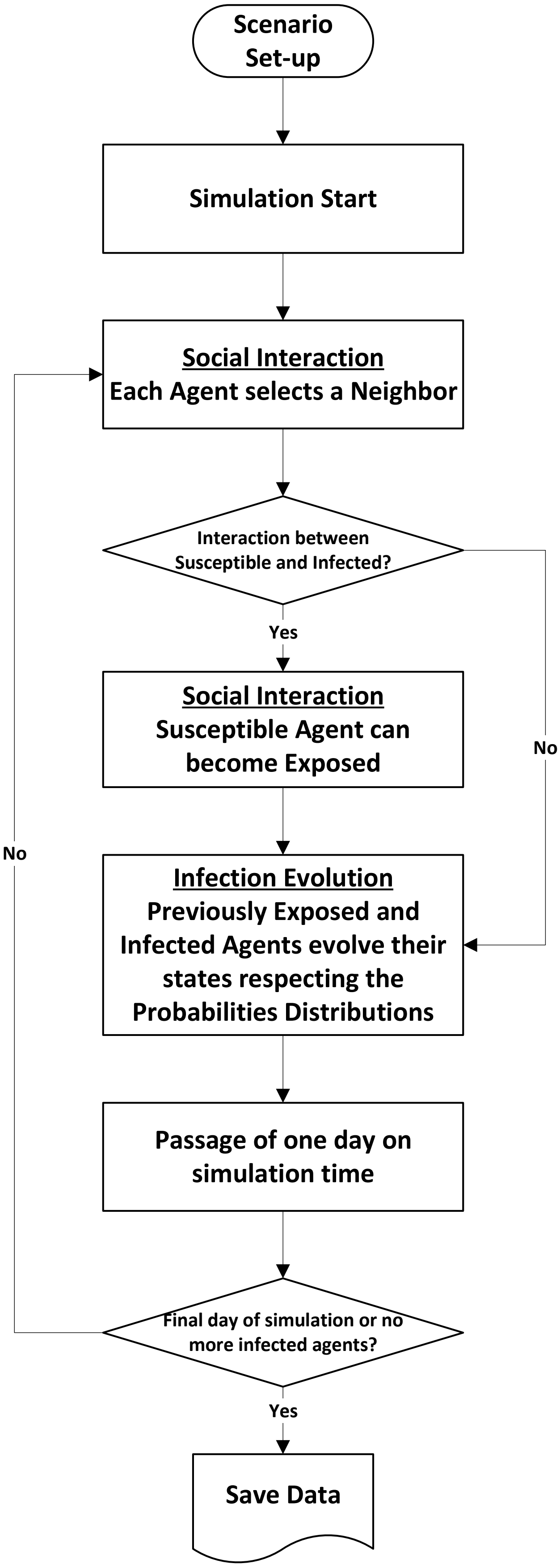


Figure 11 – Step-by-step of the simulation process applied to the illustrative example

distributions of state transitioning are set up.

In the social interaction model definition, we should set parameters related to the population and the network connecting its individuals. We must define the number of agents and their initial state, which can be drawn by randomly selecting agents based on given percentages of the population that start at each state. We can also, for instance, select part of the individuals to represent careless people, i.e., people with a higher probability of infecting/getting infected during social interaction.

Regarding the network governing social interactions, we should define: if it is static or dynamic; the number of clusters; the edges, representing possible social interactions, both inside and outside the clusters.

3.3 ASSESSING THE RESUMPTION OF PRESENTIAL TEACHING ACTIVITIES

In this section, we employ the knowledge acquired from simulating people interacting in networks as an effort to create useful predictions for the spreading of SARS-CoV-2. The model for the random interactions people have each day was the same used for addressing the information spreading problem. However, we proposed a new model for depicting each person and how their behavior changes according to the stage of the infection they are currently in. The results from the combined models are promising. The findings of this work were an asset for the scientific committee from the Federal University of Santa Catarina (UFSC). We carry a specific study to evaluate the impact of resuming presential activities in the University, the results achieved are published in an article at the website of Covid-19-SC Group (SALEM; MORENO, 2020).

The objective of carried simulations was to analyze the impact of resuming presential teaching activities at UFSC. For this purpose, we used a non-deterministic simulation model, grounded on the interaction of connected agents in a network. Each connection representing a possible (direct or indirect) contact between two persons, which may lead to an infection by SARS-CoV-2.

We considered a scenario with a single building from UFSC, assuming that this could be scaled to the total number of students without significant losses. Besides that, we also assumed that part of the students would be resistant or careless to prevention measures and, for that reason, they would have greater probability of being infected. We analyzed the effect of limiting the number of students per classroom, taking into account that this would also reduce the number of interactions per day for each agent. Finally, due to the probabilistic nature of the model, we run 100 simulations in each scenario to obtain more meaningful information from the results.

The explanation for the simulation model is divided in two parts: the model for the agents, which defines how each person acts “internally”, i.e., how they pass from one state to the other; and the model for the network, which indicates what are the possible interaction between the different agents in the simulation. Following, both components will be detailed.

3.3.1 Model for the infection evolution

In essence, each agent is defined as a finite state machine, i.e., they can be at only one state per time, and they can perform transitions from one state to another if a given pre-established condition is fulfilled. Figure 12 depicts the state machine chosen to represent the agents in the present analysis.

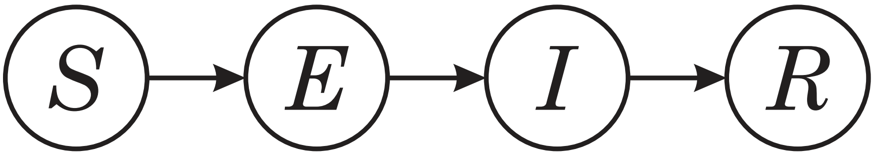


Figure 12 – State machine for the individuals, where: S → Susceptible; E → Exposed; I → Infected; R → Removed.

This states are equivalents to the compartments considered in SEIR epidemiological model, which is widely used in the literature. However, it is valid to describe their meaning to clear any question related to the simulations shown here.

- [S] Susceptible: A person that was never infected by the virus, therefore, he/she can pass to the Exposed state when in contact with an Infected person;
- [E] Exposed: A person that have been in contact with an Infected and caught the virus, but his/hers viral load is insufficient to transmit it for other persons;
- [I] Infected: A person that caught the virus some time ago and is able to infect Susceptible persons, if a contact occur;
- [R] Removed: A person that caught the virus and already passed the period in which he/she could infect other persons.

The state [R] may require further explanation. The term *Removed* was chosen because this state could represent many things, such as: a person that kept in quarantine after symptoms onset; a person that was hospitalized due to more severe symptoms; a person that recovered fully from the infection; or a person that died from complications. In short, it represents any causes that lead to the person not being able to infect others anymore.

It is also valid to emphasizes that in the present simulations, people in [R] state are not factually removed from the system. This means that, as the number of agents in [R] state rises, the probability of an agent in [S] state getting infected drops. This is due to the fact that “removed” nodes keep on being randomly picked for interacting. However, when this occur, they do not interfere in the state of others. If we do not carry simulations in this way, it would be more reasonable that the number of interactions per day dropped as the number of people to interact with also shrinks.

The transition triggers defined for the state machine in Figure 12 are non-deterministic events. The first transition, from [S] to [E], occurs when the agent in [S] randomly choose to interact with another that is in [I]. And, given that this contact occurred, there is still a probability of the agent in [S] does not get infected.

The other two transitions, from [E] to [I] and from [I] to [R], happens after a certain number of days have passed since the last transition. This means that there is a transition probability for each given day, counting from the day the agent entered the current state. Figures 13 and 14 depicts that concept. It is worth emphasizing that these probabilities accumulate. For instance, given that, in the fourth day since last transition, the agent still not passed from [E] to [I] the probability of that occurring in that day is 1 (that is, it will always occur in these circumstances).

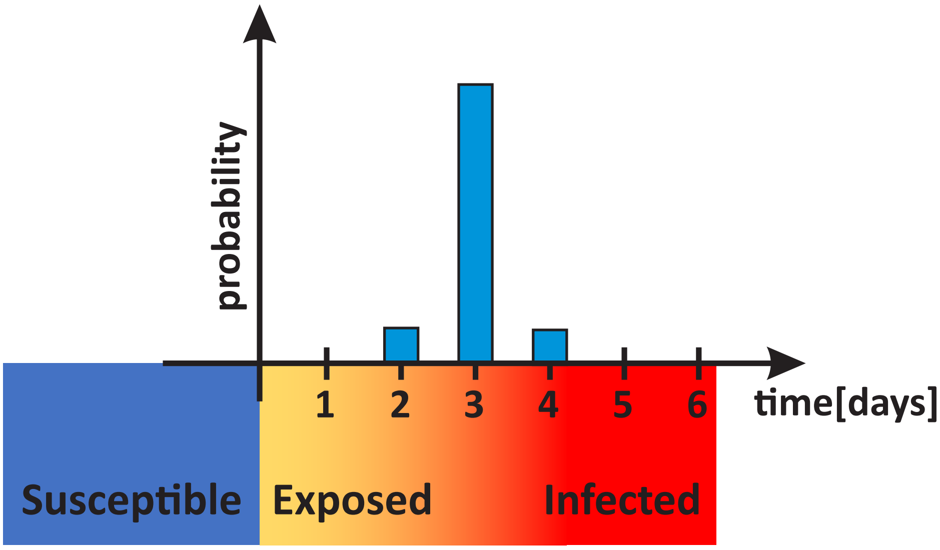


Figure 13 – Transition probability from exposed [E] to infected [I] as a function of the number of days since the transition from susceptible [S].

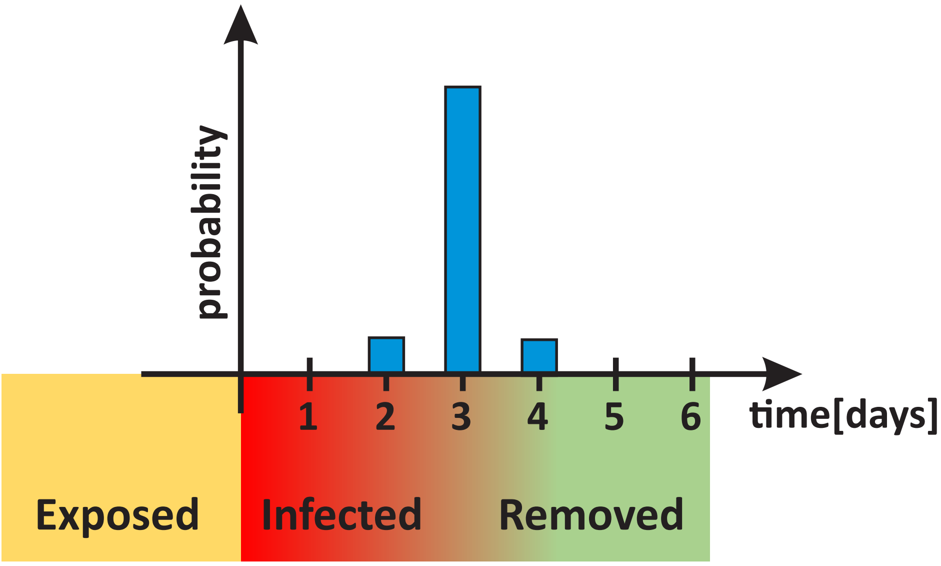


Figure 14 – Transition probability from infected [I] to removed [R] as a function of the number of days since the transition from exposed [E].

3.3.2 Model for the social interactions

In these simulations, a single building at UFSC is considered, being composed of 52 classrooms. The number of students per classroom will be given when presenting the simulation scenarios. Hence, it is assumed that each classroom is a full-connected network, i.e., a contact may occur between any members of this cluster. This setting is shown in Figure 15, with 20 students in each classroom.

The number of classrooms and students were chosen based on the structure of buildings present at our University. The assumption of a complete graph for the contacts inside the classroom is based on the fact that all students will be sharing the same closed space, which allows interactions, as well as infections, to occur between any pair of students.

Besides the contacts in the classroom, random encounters outside the cluster are considered. This latter contacts are referred in the text as “hallway contacts” for simplicity. However, they may occur in bathrooms or any other environments shared between clusters. For these hallway contacts, at each day, a Watts-Strogatz graph is generated for all the students in the building, as exemplified in Figure 16. The Watts-Strogatz graphs were generated with a mean-degree of 6 and a re-wire probability of 0.25.

Therefore, at each day, a given number of interactions happens inside the classroom and another number of contacts occur in the “hallways”. The interleaving of these two kinds of interaction is given in a random manner.

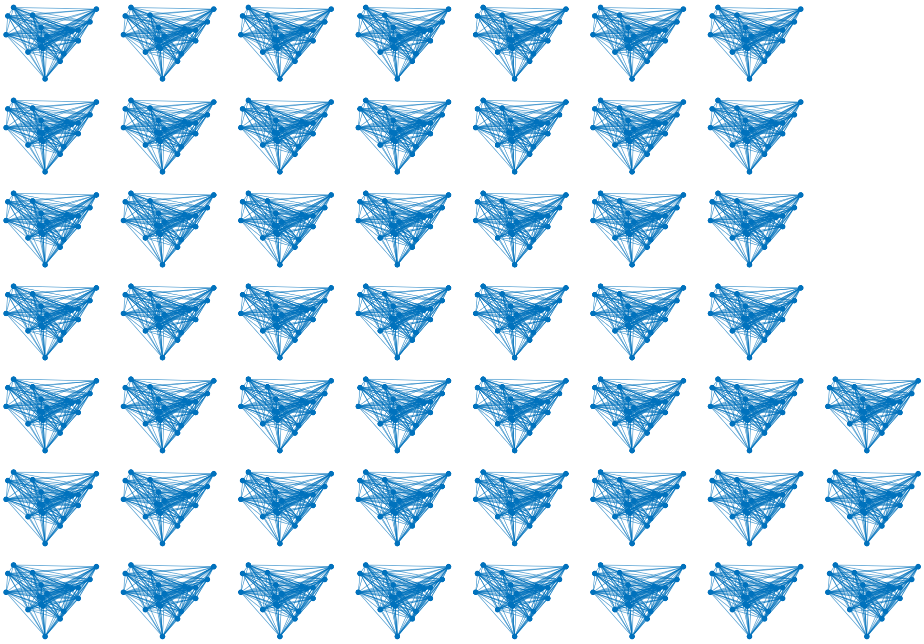


Figure 15 – Graph representing the isolated clusters (classrooms).

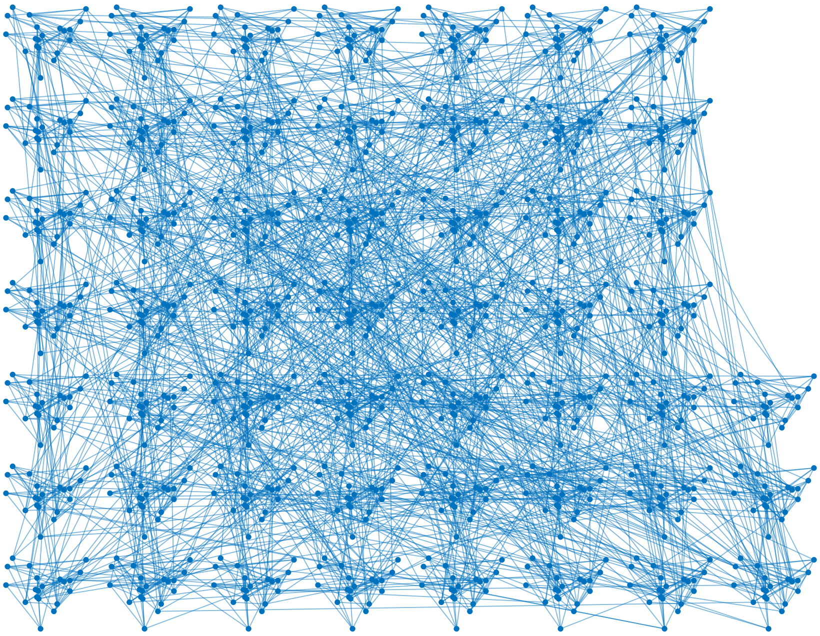


Figure 16 – Example of graph representing the random interactions between all nodes.

3.3.3 Simulations and Discussion

We considered three different scenarios for the simulations. In all of them, given that an potentially infectious contact happened (that is, between a person in [S] with another in [I]), the infection probability is defined as 0.03 (3%). In case the person is "careless", this probability rises to 0.15 (15%).

Also, in all scenarios, 5% of the total population is randomly picked to be "careless", rounding down the result to the nearer integer. The number of individuals initially in state [I] is 0.5% of the whole population, rounding up the result to the nearer integer. Individuals that does not start in state [I] will be in state [S].

Hence, what differentiates the three scenarios is the variation of the following parameters:

- Number of persons per cluster;
- Number of contacts inside the cluster (classroom) per day;
- Number of hallway contacts per day.

The name chosen to represent each scenario indicates the value set for each of these parameter in the same order that they are presented. Since the number of classrooms (clusters) is fixed, varying the number of persons per cluster will directly affect the size of the population in each scenario. Therefore, the first scenario, called **20-10-6**, will have 1040 agents in total. The second scenario, **15-7-4**, will have 780 agents. And the third one, **10-5-4**, a total population of 520.

As previously mentioned, we carried 100 simulations for each scenario. All simulations considered an interval of 30 days, such that the data referring to the 31st day is presented to depict the situation at the end of the 30th day. That is the same as saying, the final state of each day is the initial state to the following day.

Figures 17 to 19 show the results for scenario **20-10-6**. Figure 17 shows the mean values for the 100 simulations, along with the corresponding standard deviation.Figures 18 and 19 are stacked bar plots depicting only the average values. The latter, Figure 19, omits the number of persons in state [S] for better visualizing the total of persons that caught the virus in this scenario. Besides that, the latter plot present the normalized values in relation to the total population for the scenario.

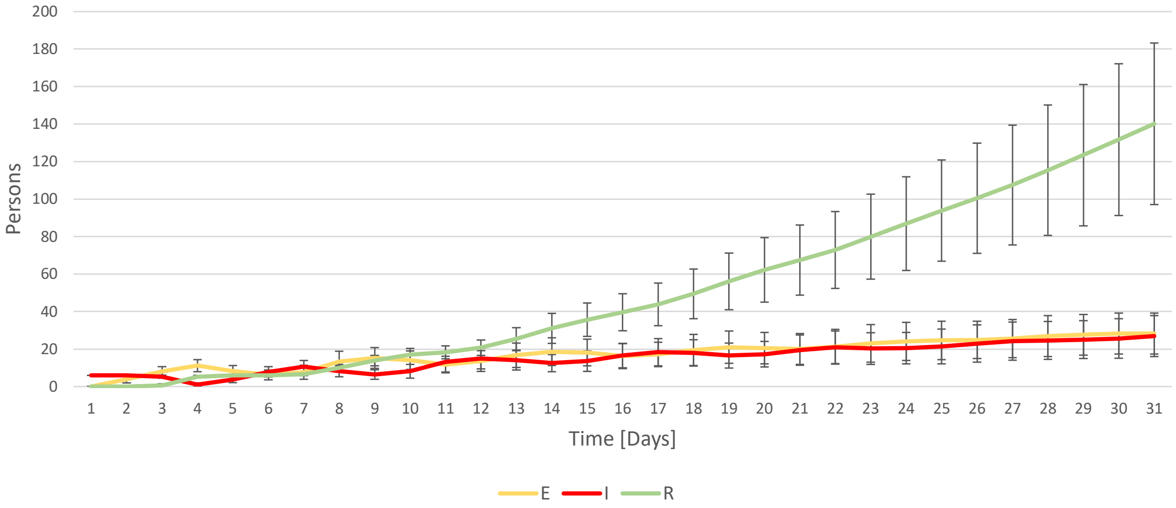


Figure 17 – Mean and standard deviation for the spreading in the **20-10-6** scenario.

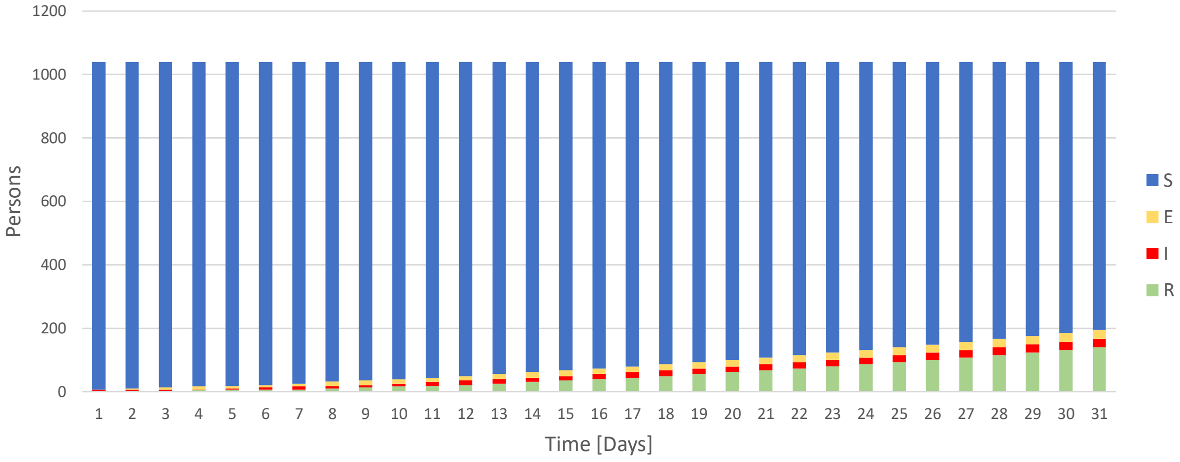


Figure 18 – Stacked bars plot for the average number of persons in each state, for the **20-10-6** scenario.

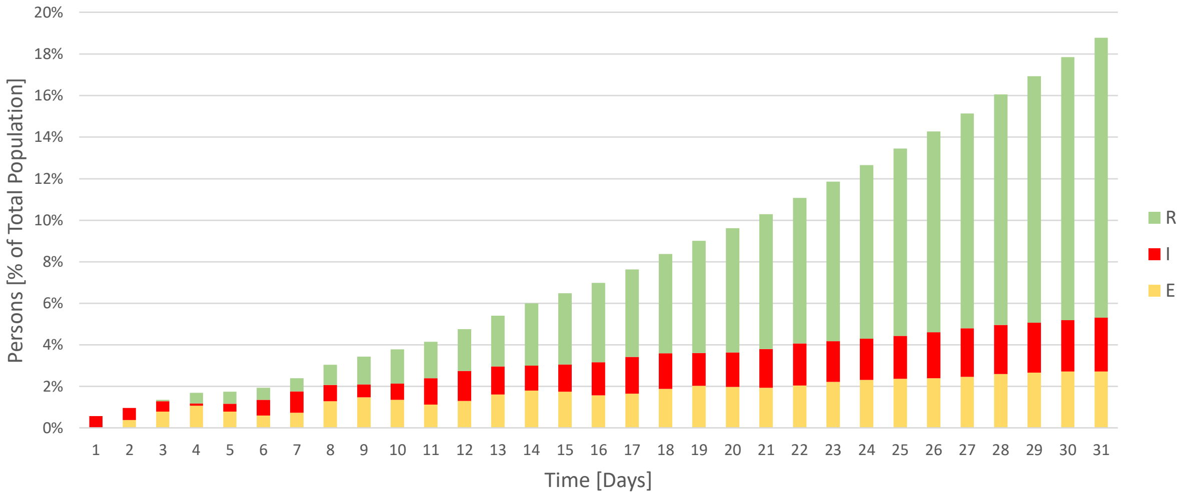


Figure 19 – Stacked bars plot for the average percentage of people that caught the virus, for the **20-10-6** scenario.

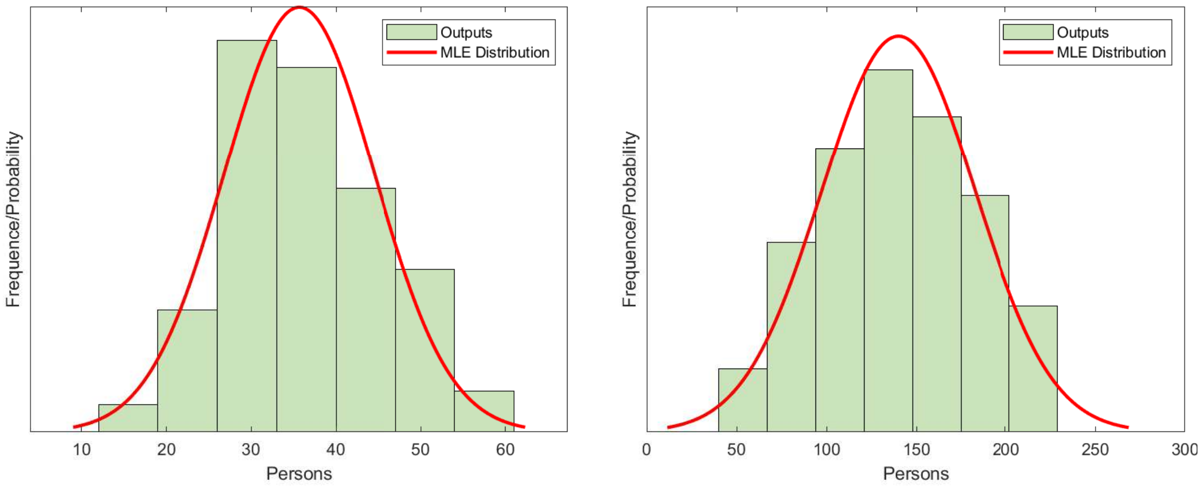
Since presenting the statistical analysis for every time step would be too extensive, we selected two days to illustrate the behavior of these random variables. We elected a point in the middle ($k = 15$) and at the end ($k = 31$) the simulations. In Table 1, we present the analysis for scenario **20-10-6**, where μ is the mean, CI (95%) is the 95% confidence interval for μ, σ is the standard deviation, and $\sigma_n = \sigma/\mu$ is the normalized standard deviation.

From the values in Table 1, we can observe that the data variance grows as the simulation days increase. This behavior is expected since the random variables always start from the same initial conditions and any deviation between two different runs of the simulation will accumulate over time.

Table 1 – Statistical analysis for scenario **20-10-6**

	μ	CI (95%)	σ	σ_n
$S_i(15)$	972.6	± 3.7	18.6	0.02
$S_i(31)$	844.7	± 12.1	61.2	0.07
$E_i(15)$	18.2	± 1.4	7.1	0.39
$E_i(31)$	28.3	± 2.2	10.9	0.39
$I_i(15)$	13.6	± 1.1	5.4	0.40
$I_i(31)$	27.0	± 2.2	10.9	0.40
$R_i(15)$	35.7	± 1.8	8.9	0.25
$R_i(31)$	140.1	± 8.5	43.1	0.31

We have run a Maximum Likelihood Estimator (MLE) for assessing if the simulations' outputs could be fitted by a normal distribution. In Figure 20, we present the results of this process for the random variable $R_i(k)$ on the same days of the previous analyses ($k = \{15, 31\}$), still for the **20-10-6** scenario. The purpose of this analysis is to illustrate that the outputs approximate a normal curve. Hence, we will not extend this representation for every random variable, on each day, and at each scenario.



(a) Day $k = 15$ (b) Day $k = 31$

Figure 20 – Histograms of values for the random variable $R_i(k)$ on different replications of the simulations for scenario **20-10-6**

Figures 21 to 23 show the results for the **15-7-4** scenario. They are presented in the same order and meaning of the previous scenario.

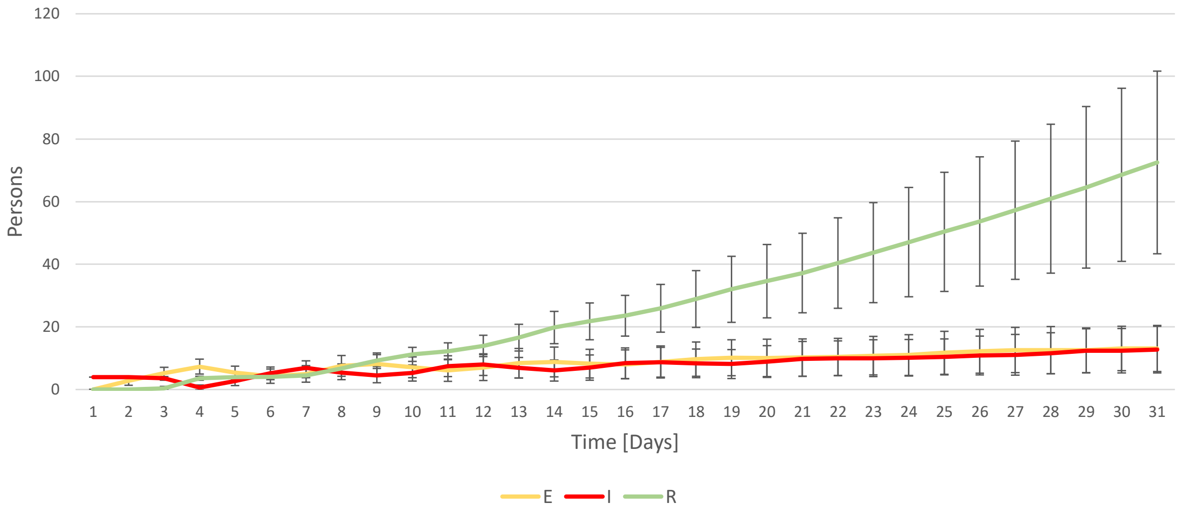


Figure 21 – Mean and standard deviation for the spreading in the **15-7-4** scenario.

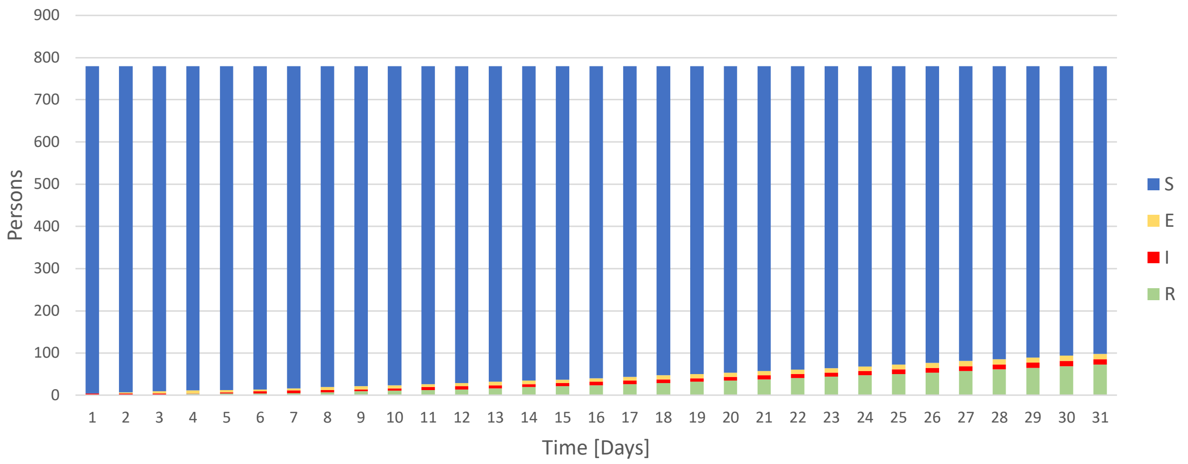


Figure 22 – Stacked bars plot for the average number of persons in each state, for the **15-7-4** scenario.

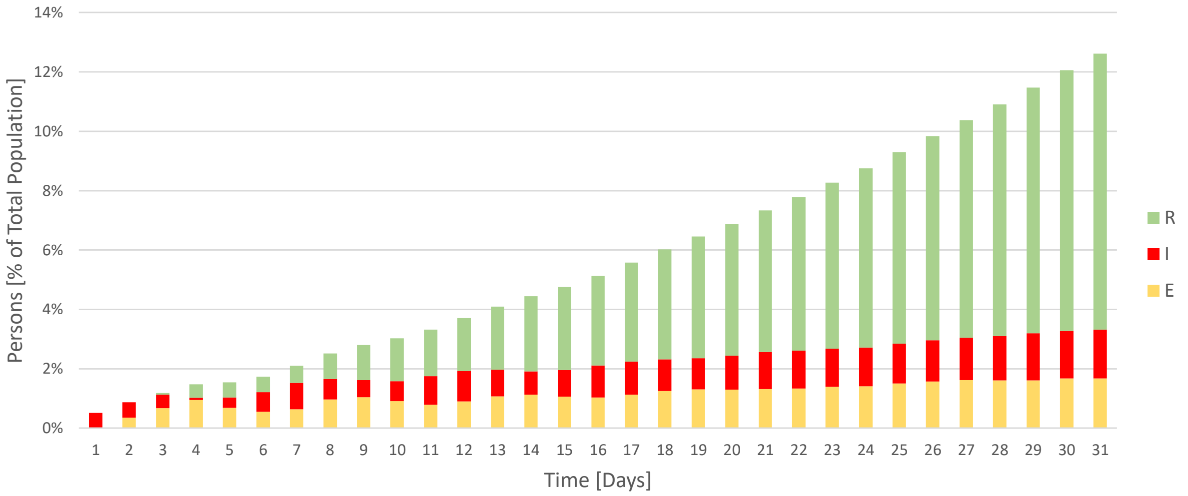


Figure 23 – Stacked bars plot for the average percentage of people that caught the virus, for the **15-7-4** scenario.

We present the statistical analysis for the **15-7-4** scenario in Table 2. The interpretation of these measurements is similar to the previous case. However, we can notice that there was an increase in the dispersion of the results, which is clearer when analyzing the normalized standard deviation, σ_n.

Figures 24 to 26 show the results for the **10-5-4** scenario. They are presented in the same order and meaning of the previous scenario.

Table 2 – Statistical analysis for scenario **15-7-4**

	μ	CI (95%)	σ	σ_n
$S_i(15)$	742.9	± 2.5	12.7	0.02
$S_i(31)$	681.6	± 8.3	41.6	0.06
$E_i(15)$	8.3	± 0.9	4.6	0.55
$E_i(31)$	13.12	± 1.4	7.3	0.56
$I_i(15)$	7.0	± 0.8	4.0	0.57
$I_i(31)$	12.8	± 1.5	7.4	0.58
$R_i(15)$	21.8	± 1.2	5.9	0.27
$R_i(31)$	72.5	± 5.8	29.2	0.40

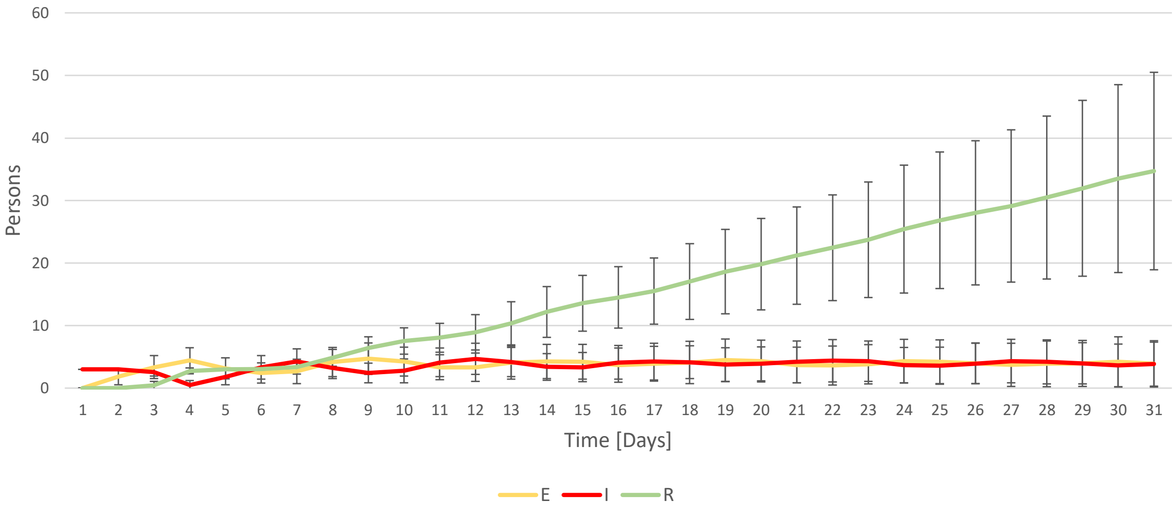


Figure 24 – Mean and standard deviation for the spreading in the **10-5-4** scenario.

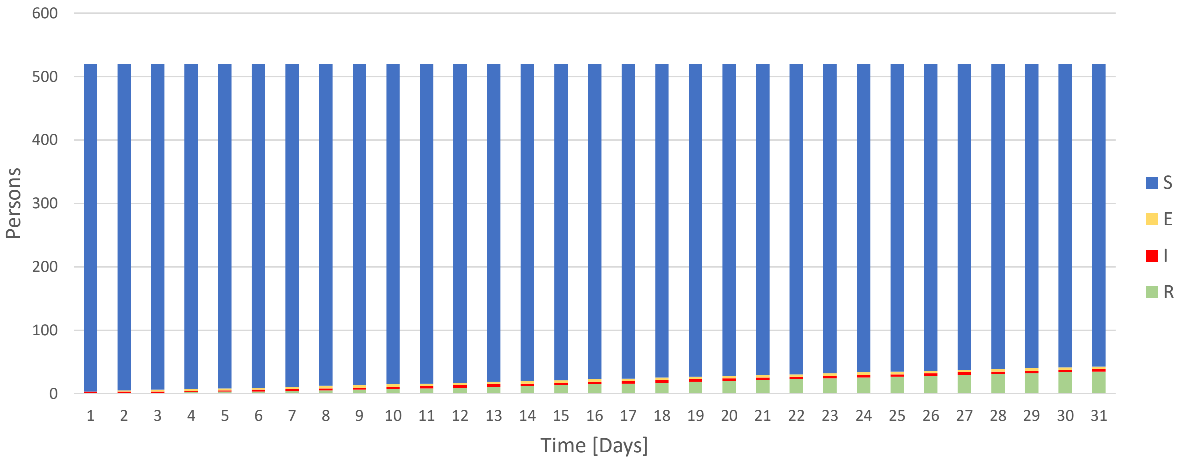


Figure 25 – Stacked bars plot for the average number of persons in each state, for the **10-5-4** scenario.

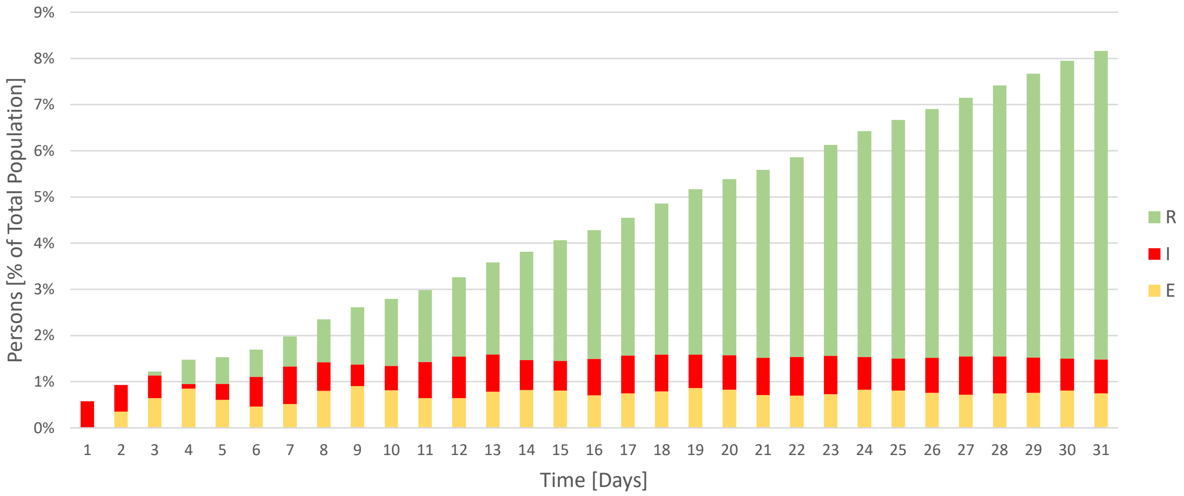


Figure 26 – Stacked bars plot for the average percentage of people that caught the virus, for the **10-5-4** scenario.

Furthermore, in Table 3, we summarize the data from the statistical analysis over the **10-5-4** scenario. The interpretation of these measurements is similar to the previous scenarios. Once again, we can observe an increase in the standard deviation of most of the random variables. This can indicate a general behavior brought to sight by our simulations, which is the increase in results variance in inverse proportion to the population size.

Table 3 – Statistical analysis for scenario **10-5-4**

	μ	CI (95%)	σ	σ_n
$S_i(15)$	498.9	$\pm$ 1.5	7.7	0.01
$S_i(31)$	477.6	$\pm$ 4.2	21.2	0.04
$E_i(15)$	4.2	$\pm$ 0.5	2.8	0.66
$E_i(31)$	3.9	$\pm$ 0.7	3.7	0.96
$I_i(15)$	3.3	$\pm$ 0.5	2.3	0.69
$I_i(31)$	3.8	$\pm$ 0.7	3.5	0.92
$R_i(15)$	13.6	$\pm$ 0.9	4.5	0.33
$R_i(31)$	34.7	$\pm$ 3.1	15.8	0.45

Even considering a low infection probability, resuming presential activities at UFSC could leave to a new wave of infections in the city. However, the simulations indicate that a restriction in the number of students per classroom could be effective in counteracting contagion. In relative terms, classrooms with 20 students would leave to a infection of about 19% of the whole population at the end of 30 days. In classrooms with only 10 students the infection would drop to little more than 8% of the total population. It is worthy to remember that both scenarios start with 0.5% of infected agents.

Even that this restriction have a positive impact, it is not sufficient for definitively holding the virus spread in campus. It would be necessary some measure to reduce the time a person is actively infecting others (state [**I**]). However, testing all students every day, or even every two days, is unfeasible.

Hence, a high risk is verified in the possibility of resuming fully presential activities in a near future. The shrinking of this risk is associated to restricting the number of contacts, by reducing either the number of students per classroom or the time that students remain in the University. It is important to emphasize that this study did not take into account the contamination in the displacement towards the University, neither the interaction with models of propagation in the city as whole. Both factors would impact significantly in worsening the results presented here.

3.4 PREDICTING THE VIRUS SPREAD IN A CITY

The same principles that grounded the simulations in previous section can be used to obtain results that provide more information to specific problems. For instance, the number of people needing medical attention at the same time, and those who will be taken to Intensive Care Units (**ICUs**) are critical knowledge for the city administration. Being able to predict this allows a better resource allocation, which impacts directly in preventing deaths.

3.4.1 Model for the infection evolution

The concept of transitions with time-dependent probabilities presented in the previous section is also employed in this case. However, the probabilities will not be explicitly depicted (as they were in Figures 13 and 14) for two reasons: there are many transitions in this model; and the probability values are defined without a strong foundation due to the disease novelty.

The point is that the model is flexible enough to easily adapt for new estimates of the probabilities as more data is collected and the scientific community advances toward a better understanding of the disease. The states and transitions considered for the simulations shown ahead are depicted in Figure 27.

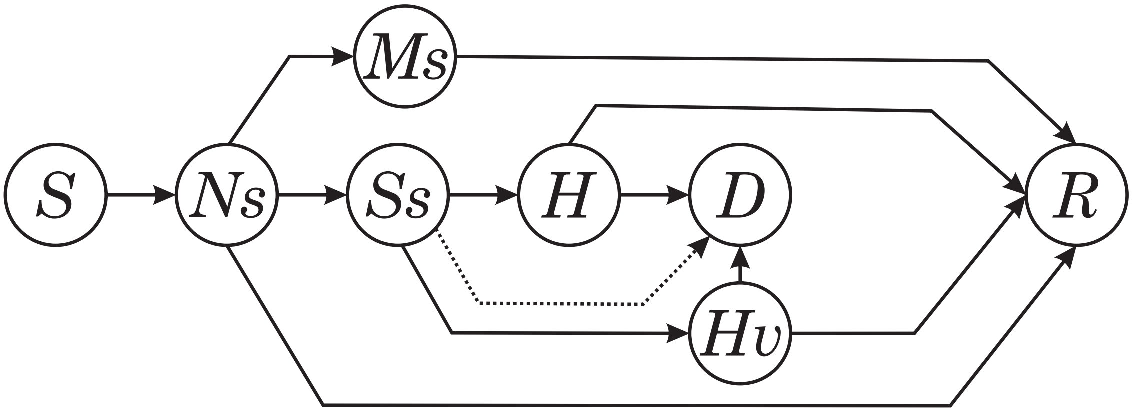


Figure 27 – State machine for the individuals, where: S → Susceptible; Ns → No symptoms; Ms → Moderate symptoms; Ss → Severe symptoms; H → Hospitalized; Hv → Hospitalized using ventilator; R → Recovered; D → Dead.

The states considered for this model can be viewed as subdivisions from the ones in SEIR-based model presented previously. This division into more specific groups allows the agents to have diverse behaviors according to their current health condition. The meaning of each state in the simulation is given as follows.

- [S] Susceptible: A person that never got infected by the virus, therefore, he/she can pass to state [Ns] when in contact with an infected person (i.e., in state [Ns], [Ms], [Ss], [H] or [Hv]);
- [Ns] No symptoms: A person that already caught the virus but do not present symptoms, his/her condition can evolve to state [Ms], [Ss] or even [R];
- [Ms] Moderate symptoms: A person that presents mild to moderate symptoms, he/she will pass to state [R] without needing to be hospitalized;
- [Ss] Severe symptoms: A person presenting symptoms that require medical attention, he/she can pass to state [H] or [Hv];
- [H] Hospitalized: A person admitted into a hospital, he/she can either recover (state [R]) or die (state [D]);
- [Hv] Hospitalized using ventilator: A person in critical situation, requiring the use of a ventilator to keep she/him alive;
- [R] Recovered: A person that caught the virus but no longer can infect others;
- [D] Dead: A person that deceased due to disease complications.

The dashed transition from [Ss] to [D] is only used for scenarios considering the health system overload. More specifically, it represents the sad situation when a person requires a ventilator and there are none available. Hence, instead of going to an ICU bed and being intubated (state [Hv]), the patient dies (state [D]).

This model enables several possibilities of policies evaluation. For instance, one may want to compare the effect of a complete lockdown versus quarantining only people showing symptoms. This could be made by constraining (or even suppressing completely) the number of daily contacts that an agent will have while in a given state. However, comparisons in this directions are hitherto unexplored.

It is worth pointing out that some states in this model can lead to more than one other state (e.g., [Ns] can transition either to [Ms], [Ss] or [R]). In these cases, the maximum value for the accumulated probability of each transition should not be 1, but rather the portion of cases that evolve to each corresponding states. As an example, in the simulation illustrated in Figure 33, it is assumed that from all infected persons in state [Ns], 55% will display moderate symptoms (state [Ms]), 15% will have severe symptoms (state [Ss]), and 30% will remain asymptomatic until recovery (state [R]). This numbers are constantly being updated. Currently, more accurate values could be drawn from the work of Stokes et al. (2020).

3.4.2 Model for the social interactions

In the simulations presented here, we defined the network topology as clusters representing the neighborhoods, in which a Watts-Strogatz network defines how people are related. Besides that, a certain number of random temporary connections between clusters is created each day.

Figure 28 depicts the initial graph for the network in the simulation. It is an abstraction of the city of Florianópolis (Santa Catarina/Brazil) scaled down by a factor of 100. Hence, a total population of 4818 agents is considered, equally divided into 66 clusters (neighborhoods). At the beginning of each day, 48 random links (1% of the population) are created between different clusters. At the end of each day, these links are removed.

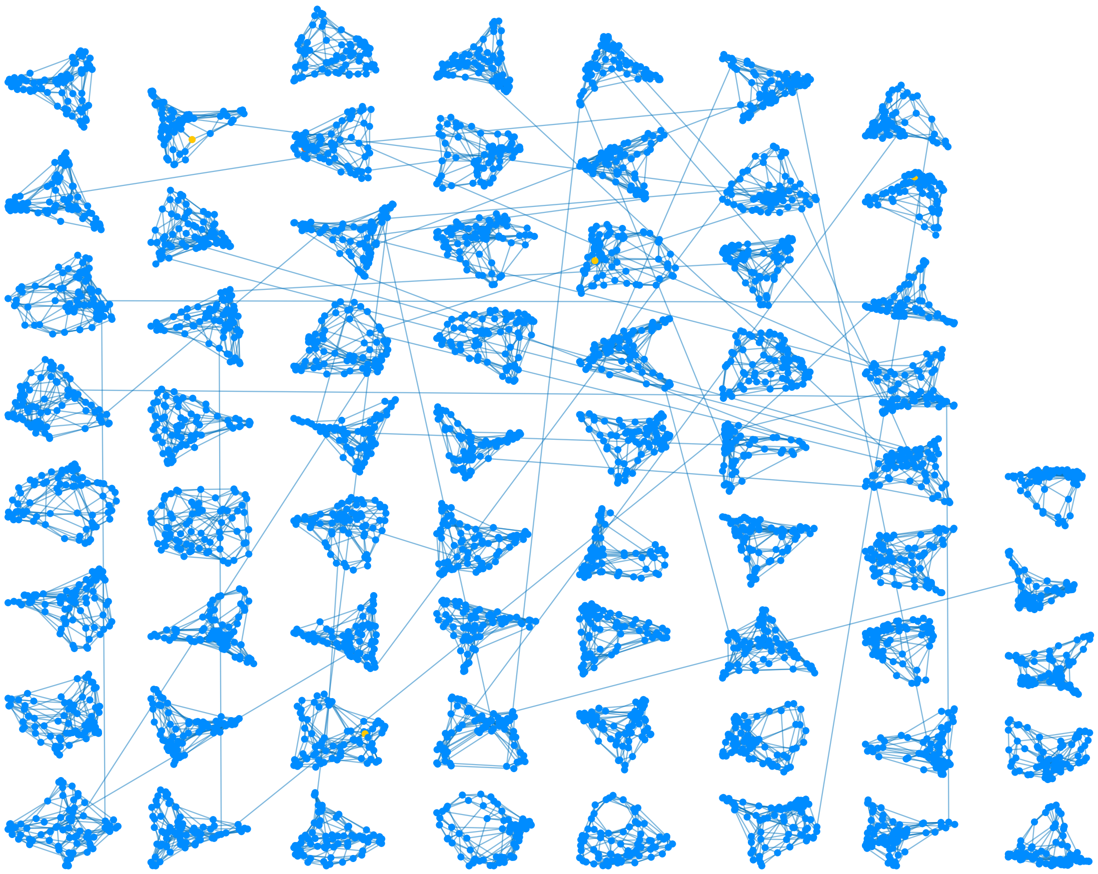


Figure 28 – Example of a graph at the beginning of the simulation.

To initialize the simulation, 4 agents are randomly drawn from the entire population to start in state [Ns] (represented by yellow nodes in the graph). Similarly, 1 agent is selected to start in state [Ms] (represented by the orange node in the graph). In Figure 29, we present the QR codes that leads to videos showing two illustrative scenarios of this simulation setup.

(a) No limiting ventilators, also accessible in: https://youtu.be/gjyyy1ox-rk

(b) Limiting ventilators, also accessible in: https://youtu.be/U-9_F5vx2Ps

Figure 29 – QR Codes directing for animations showing the Covid spreading evolution according to our model in two different scenarios.

3.4.3 Simulations and Discussion

The simulations shown here seek to illustrate the model's potential and flexibility. Initially, a comparison is made between two scenarios that differ from each other by the number of available ventilators. In both scenarios, 100 simulations were carried to obtain more meaningful results, reducing the effect of randomness in the analysis.

The first scenario in this comparison considers unlimited ventilators, i.e., every agent needing a ventilator will successfully pass to state [Hv]. Figure 30 shows the mean values on the 100 simulations for this case, along with the corresponding standard deviations. In this scenario, Infection Fatality Rate (IFR) is 1.32%. IFR is defined as the number of dead persons divided by the total number of persons that caught the virus. At the epidemics end, the total number of persons that caught the virus is simply the sum of agents in states [R] and [D].

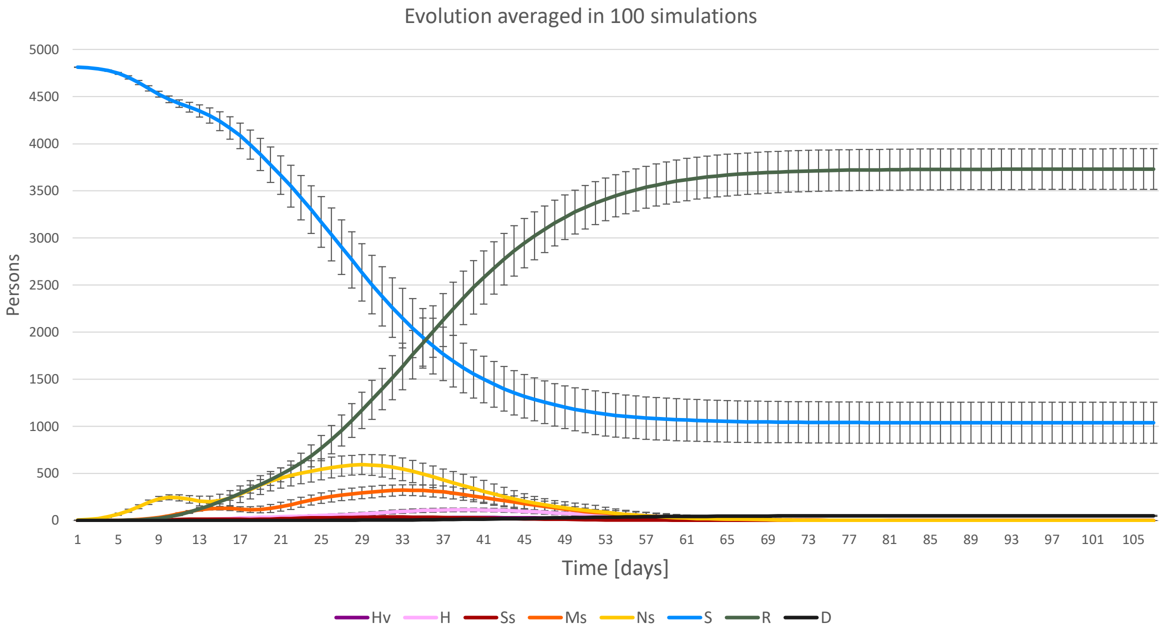


Figure 30 – Evolution of cases averaged over 100 simulations, not considering the health system overload.

The second scenario considers that only 4 ventilators are available. This may sounds too little, but it is, in fact, the number of ventilators available in Florianópolis after applying the same scale factor used in the total population. Figure 31 shows the mean values on the 100 simulations for this case, along with the corresponding standard deviations. Here, the IFR rises to 1.99%. This represents a increase of 50.8% in the number of deaths.

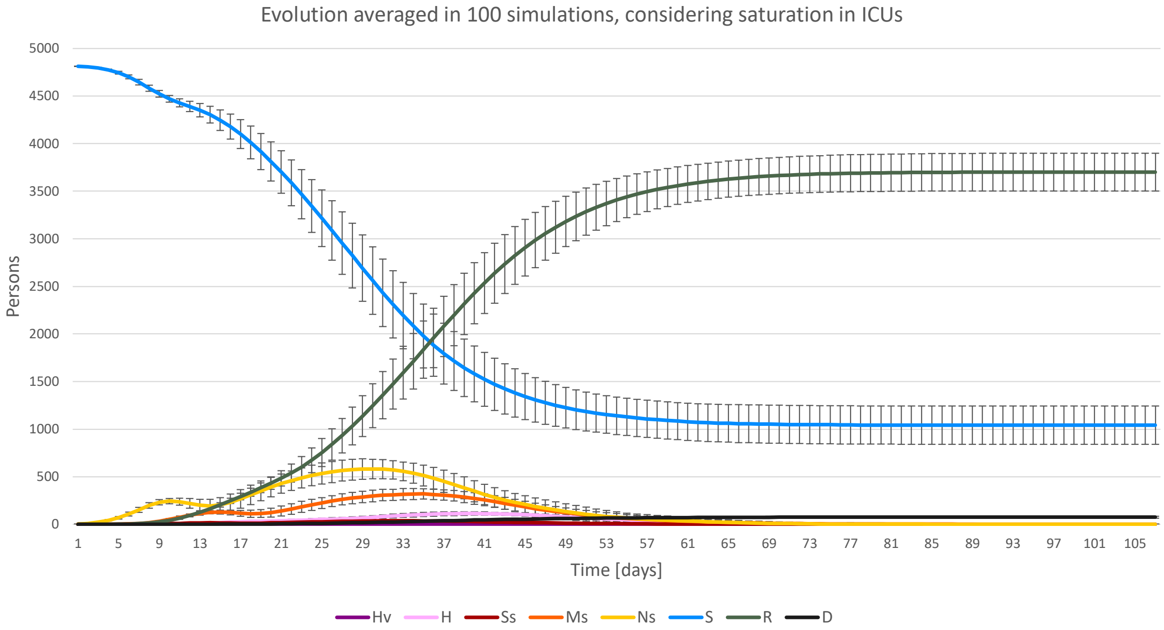


Figure 31 – Evolution of cases averaged over 100 simulations, considering the health system overload.

To depict the number of people who would die by the lack of ventilators, Figure 32 shows the evolution of how many agents were in [Hv] state on the first scenario, alongside with the ventilator-equipped ICUs capacity considered in the second scenario.

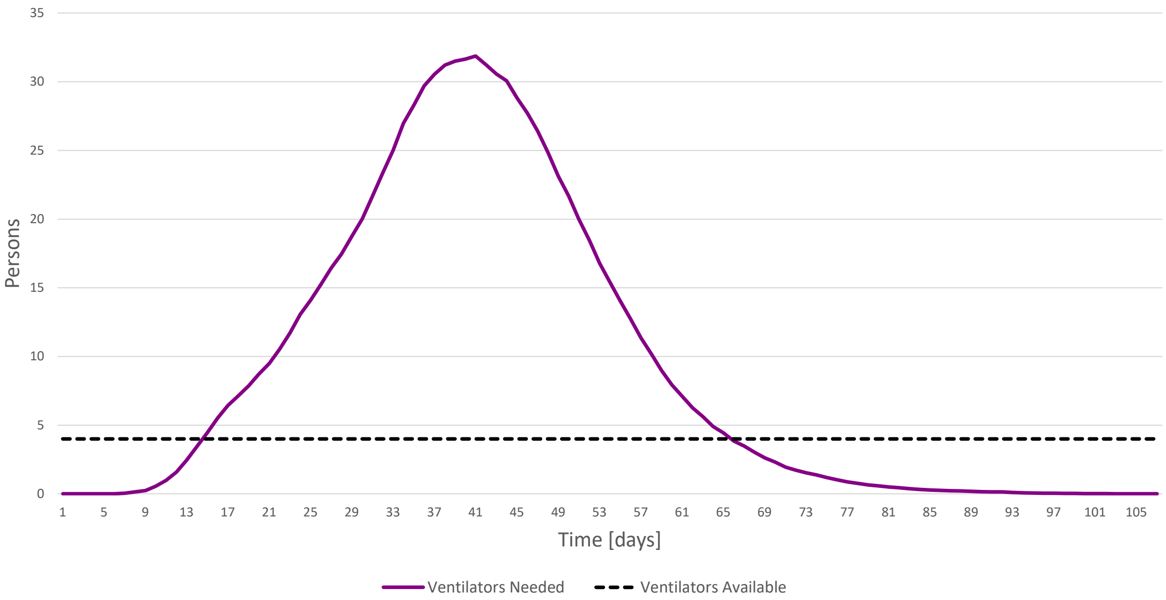


Figure 32 – Comparison of ventilators demanded versus available (averaged over 100 simulations).

In absolute numbers, the average of deaths in the first scenario is 49.99, while it is 75.1 in the second one. Considering the scale factor employed, this means 2511 more deaths in only one city.

At last, we present another computational experiment. Here, the number of simulation rises to 1000 for evaluating the effect in standard deviations. Also, the transition probabilities are changed to illustrate the impact of these parameters in the overall number of agents in each state. The results from this experiment are shown in Figure 33.

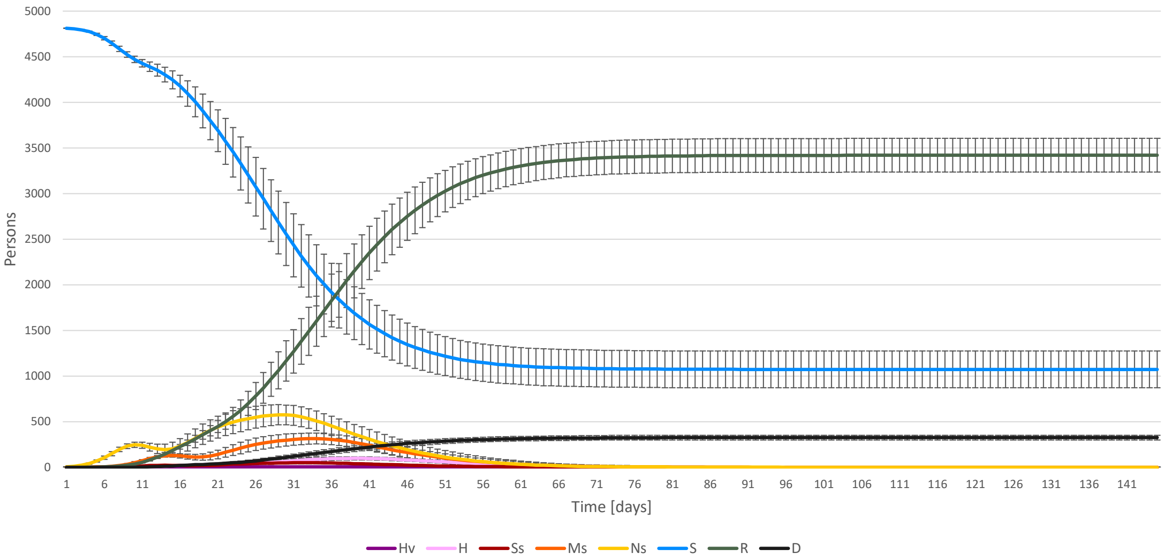


Figure 33 – Evolution of cases averaged over 1000 simulations.

It is noticeable that the standard deviations did not changed much in relation to the last experiment. Hence, one can assume that 100 simulations is enough to obtain reliable information. Since simulations with many agents take a long time to be computed, this is an important result to make the experiments feasible.

Also, the number of deaths highlights the impact of changing the transition probabilities. This feature can be used to validate the model. Using parameters from another virus or contagious disease that have more reliable and established data, the model can be tuned to predict its spreading and compared to historical information on the subject disease.

4 INFORMATION SPREADING

The development of applications on the web 2.0, combined with portable devices and Artificial Intelligence (AI) algorithms, the first Online Social Networks (OSNs) applications designed to get people in contact and make friends, turned into social networks made up of over 2 billion users. This combination of communication networks, portable devices, and AI has irreversibly changed the way people interact and make decisions (VAN DIJCK, 2013).

This impact from OSNs brought the attention of many researchers, especially in the community of Systems & Control, due to possibilities of applying well-known tools of representation, such as graph theory, and the utilization of dynamical systems to model the diffusion of information and the achievement of consensus. In some measure, these problems are very similar to one studied in cooperative mobile robotics and other applications (PROSKURNIKOV; TEMPO, 2017; BULLO, 2019).

Besides the developments in control and modeling of social networks, many philosophers, sociologists, and political scientists are addressing these new phenomena that arouse in social behavior by the massive utilization of social networks (HAN, 2017). Not only to understand the newly emerging collective behaviors but to foresee possible consequences and ethical dilemmas in the usage and manipulation of social networks (VALLOR, 2016; HOWARD et al., 2018).

Despite the vast possibilities and innovations generated by OSNs, many of them being an essential tool for our everyday life; like any other disruptive technology, society must be aware and discuss potential undesirable side effects. In the last years, many disturbing effects of the abuse or misuse from OSNs have been reported in media. Manipulation of public opinion in elections by massive robotic nodes, mob attacking innocent people, non-precedent diffusion of fake news, bullying among teenagers, profiling of users overruling privacy, and many others.

The first works considering ethical questions on OSNs were more related to the effects of the overexposure and privacy breaches. However, with the use of AI to make profiles from users on social networks, more attention has been given to the possible manipulations of will, desires, and free choices. More recently, the use of apps to share messages, as a tool for disseminating fake news and influencing elections, was at the center of the debate of the misuse of OSNs. Following previous works on the modeling of opinion diffusion in social networks and trying to consider ethical questions, one of the subjects addressed in this document is the analysis of the effects of information dissemination.

The first internet-based social networks were launched at the end of the 1990s, allowing people to create profiles and list friends and classmates (BOYD, D. M.; ELLISON, 2007). After the proliferation of many platforms, the instant messaging services increase the usage of OSNs, but the great step was given with the applications that enable sharing of content, turning these platforms into large corporations, and the applications into techno-cultural devices (VAN DIJCK, 2013).

New manifestations emerged with the use of OSNs, such as the dissemination decentralization of cultural products, as well as news and opinions through blogs, and the use of social networking platforms, among others. These manifestations soon attracted the attention of scholars, trying to analyze possible ethical and social impacts, especially from social networking platforms. Many philosophers of technology, such as Albert Borgmann and Hubert Dreyfus, pointed out some possible risks to human relations from the overexposure in social networking (VALLOR, 2016). However, due to the continued development of new applications, the widespread popularization of mobile phones, and the use of AI, the previous analyses were overwhelmed by new mass behaviors on a global scale.

Nowadays, by using shared information provided by users, applications are designed to support many activities, such as finding a job, renting a house, or even connecting citizens to run a neighborhood. One of the most critical aspects of OSNs is the free exchange of information by users in small or large groups. These new uses have changed many aspects of everyday life due to new social and political phenomena. According to many studies, the utilization of OSN changed the political space. The democratization of access to media, information production or consumption, and the increase of the power of mobilization were some of the major benefits.

However, the constitution of large groups to exchange ideas and information, gave rise to a phenomenon of polarization (HAN, 2017) and a "*tendency to reward virality over veracity may harm information quality and democratic discourse*" (NEUDERT; MARCHAL, 2019).

OSNs have been massively used, across the world, not only as a democratic tool for the dissemination and debate of ideas and proposals, but also for manipulation purposes (TARDÁGUILA; BENEVENUTO; ORTELLADO, 2018; BRADSHAW; HOWARD, Philip, 2017). Many factors contributed for increasing this type of usage, such as: lack of legislation, distributed structure, tendency to virality, and characteristics of fast and deep propagation of *fake news*, due to novelty and emotional appeal of the content (VOSOUGHI; ROY; ARAL, 2018),

This manipulation is based on the diffusion of personalized messages grounded on tools for profiling users but also with the use of robotic nodes, *social media bots*, in messaging applications (HOWARD et al., 2018; WANG, P.; ANGARITA; RENNA, 2018).

The ethical issues of these misuse of OSNs are related not only to privacy and security but with trust and social manipulation, undermining democratic processes, urging the development of policies and regulation tools (TURCULEŢ, 2014; HOWARD et al., 2018). Another important aspect is the impact of the network misinformation in health decisions, by the spreading of *fake science*, affecting the public health system (SHAO et al., 2018). This became explicit during the Covid-19 pandemics.

The proposal of strategies to prevent the misuse and manipulation of large groups in social platforms, together with the studies about social behavior, constitutes a major concern for academia, legislators, and social actors.

The study of policies to avoid the misuse of private information has received considerable attention, and the adoption of new legislation has been debated and approved. On the other hand, the detection of fake news web-based services and algorithms to fact-checking are becoming essential tools (ZHANG, X.; GHORBANI, 2019). Also, some measures could be taken to limit the number of members of groups in instant messaging applications and the number of times that users can forward messages.

Motivated by these concerns, we propose a model for the information spreading in OSNs. In this chapter, we present the model formulation and some extensions that considers some of the many aspects that may influence on a human's decision process over accepting (and possibly sharing) a given piece of information. We also provide an analysis of the node centrality's effect on the output of our model.

4.1 INFORMATION EXCHANGE MODEL

One of the concerns of this research is the spreading of information through Online Social Networks. So, modeling the agent consisted in describing the mechanism it uses to choose which information is accepted. Opinion dynamics and consensus were the fields that grounded this modeling. Hence, it is worth mentioning that the terms opinion and information may be used in this chapter interchangeably.

The literature on consensus brings applications that deal with continuous opinions, i.e., where the opinion of a robot on an issue belongs to the set of real numbers. For instance, this is the case for multi-robots systems handling problems such as formation control (CORREIA; MORENO, 2015), where the robots need to synchronize its velocities, and rendezvous in space (SALEM; MORENO; CASTELAN, 2018), where the robotic team decides the best point to meet.

However, when analyzing information spreading in networks, it is not appropriate to assume that the agents will perform a computation and come up with a new piece of information. Hence, the gossip model is altered to encompass the behavior of agents that only exchange information from a finite set.

From the original gossip algorithm (BOYD, S. et al., 2006), whenever an interaction occurs, the active agent will update its information as

$$x_{k+1}^i = (1-\nu^i)x_k^i + \nu^i x_k^j \ , \ i = i_k \tag{7}$$

where ν^i is a constant representing the trust of agent i in its neighbors.

The other agents (including j) persist their information to the next iteration, i.e.,

$$x^l_{k+1} = x^l_k \ , \ \forall l \neq i_k \tag{8}$$

The definition made in (6) is an important step to handle discrete information. Since the matrix of interaction probabilities now represents the trust of the agents to one another, a new assumption over the gossip model can be made.

In (7), consider that $\nu^i = 1 \ , \ \forall i$. Then, it simplifies to

$$x^i_{k+1} = x^j_k \tag{9}$$

where x^i_k is the information from agent i. This way, whenever an interaction occurs, the active agent does not compute a weighted average of its information and the one displayed by its chosen neighbor.

Instead, the active agent will replace its opinion entirely, which guarantees that the information will be only exchanged, and no new information is created within the network. This way, the present model is said to handle discrete information. Originally, the gossip algorithm did not allow an agent to interact with itself, i.e., $p_{ii} = 0 \ \forall i$. This is reasonable given how the agents updates their information in that model, but now this constraint is toppled to enable agents sticking with their actual information.

Figure 34 – QR Code directing for the animation illustrating the consensus with information from a finite set, also accessible through the link: https://youtu.be/R3nEx1tqL_I.

Figure 34 shows a QR Code for viewing an animation illustrating the information exchange according to our model. A fundamental aspect of this model is that the connections and its weights are the only factors influencing the information-exchange process. In other words, at this point, the content of the information is irrelevant.

Further, the fact that the information spreading evolution in our model depends only on the current state enable using Markov chains to describe the same dissemination process. However, if we modeled it as a Markov chain, we would have to create a different state for each combination between the agents and the pieces of information exchanged in the network, making the computational cost unfeasible with many agents and/or pieces of information.

In his thesis, Carvalho (2020) explored further this question, stating that such systems can be seen as a Markov chain with N^n states, where N is the number of unique pieces of information and n is the number of agents. Therefore, each state of this Markov chain would be an n-dimensional vector containing the current information of each node. Also, he concludes that this exact Markov chain used to describe the information spreading in small systems becomes impracticable in real-world problems, being too complex to be applied even for a few numbers of nodes.

During the development of this thesis, we were able to envision and provide models for two situations when the interactions' probabilities may change through time accordingly to factors different from the interpersonal trust linking the agents. These factors are the person's self-esteem over a given issue, which changes the probability of accepting information sent by others; and the information's reliability, which impacts the probability to someone accepting it.

4.1.1 Modelling the Effect of Self-esteem

The contribution presented in this section is the mathematical description of the relative interactions matrix, from the DeGroot-Friedkin model, for a scenario where some of the nodes are robots. The reason for choosing the DeGroot-Friedkin model is because it addresses the evolution of the self-weights across a sequence of issues being discussed by a group. These self-weights are analogous to individuals self-confidence or self-esteem.

Also, we present some simulations results to illustrate the model's behavior in two meaningful scenarios, followed by some analysis of the opinion consensus, the number of iterations, and comparisons between initial self-weights and stabilized self-weights. Moreover, we show the results with a larger group to demonstrate the model's scalability.

The DeGroot–Friedkin model, presented by Jia et al. (2015), predicts the evolution of an influence network governing the process of opinion formation, based on the relative interpersonal weights. The model predicts that the social power ranking among individuals evolves across issues and asymptotically tends to their centrality score. These analyses of issue sequences and this proposed formalization of evolution are motivated by the sociological hypothesis of reflected appraisals by Cooley (2017).

The available empirical evidence is also consistent with the assumption that individuals update their opinions as convex combinations of their own and the displayed opinions by others, based on weights that are automatically generated by individuals in their responses to the displayed opinions of others (JIA et al., 2015).

Based upon this, a social influence network can be described as a weight matrix $\mathcal{W} = [w_{ij}]$ satisfying $w_{ij} \in [0,1]$ for all i and j and $\sum_j w_{ij} = 1$ for all i (that is, W is *row-stochastic*). Each edge of this network $j \xrightarrow{w_{ij}} i$, including self-loops

$i \xrightarrow{w_{ii}} i$,represents the influence weight accorded by agent i to agent j. Representing the opinions/information of the individuals with a real-valued vector x, the classic DeGroot model (DEGROOT, 1974) was given in (5).

Sustained on empirical observations, Jia et al. (2015) combine the DeGroot model (5) for opinion dynamics, in which the influence network for a particular issue is fixed, with the formalization made by Friedkin (2011) to the evolution of interpersonal influences in an issue sequence. The resulting dynamical process is referred as the DeGroot–Friedkin model and is given as follows

$$x_{s,k+1} = \mathcal{W}_s x_{s,k} \ , \quad s = 1, 2, \dots, \quad k = 1, 2, \dots \tag{10}$$

where s indicates the issue that is being discussed and $\mathcal{W}_s = [w_{ij}(s)]$ is the issue-dependent matrix of influence weights.

From (10), it is possible to notice that the DeGroot-Friedkin model assumes that timescales for the two processes are separate: the opinion dynamics are faster than the reflected appraisal dynamics in the influence network. In other words, opinion consensus is achieved before individual self-weights are updated.

For notation simplicity, the self-weights, $w_{ii}(s)$, will be denoted as χ_s^i and they are the core of the model, since the interpersonal weights are evaluated using their values. Each weight $w_{ij}(s)$, $i \neq j$, satisfies $w_{ij}(s) = (1 - \chi_s^i)c_{ij}$, where the *relative interpersonal weights* c_{ij} are static and issue independent.

In an influence network with $n \geq 2$ agents, the self-weights are updated as $\chi_{s+1} = F(\chi_s)$, where $F : \Delta_n \to \Delta_n$ (Δ_n being the n-simplex) is a continuous map defined by

$$F(\chi) = \begin{cases} \mathbb{e}_i & , \ \chi = \mathbb{e}_i \ \forall i \\ \left(\dfrac{\sigma^1}{1-\chi^1}, \dots, \dfrac{\sigma^n}{1-\chi^n} \right)^T \Big/ \displaystyle\sum_{i=1}^{n} \dfrac{\sigma^i}{1-\chi^i} & \text{, otherwise} \end{cases} \tag{11}$$

where $\mathbb{e}_i$ is the i-th basis vector (all elements equal to 0 except the i-th, which is equal to 1), $\sigma^\top = \begin{bmatrix} \sigma^1 & \dots & \sigma^n \end{bmatrix}$ is the vector of centrality scores for each agent and it is defined as the dominant left eigenvector of the relative interaction matrix $\mathcal{C} = [c_{ij}] \in \mathbb{R}^{n \times n}$, which is row-stochastic, zero-diagonal, and irreducible.

Including agents with static self-weights

Now, we consider the DeGroot-Friedkin model with two types of agents. The first one encompass the human-agents, whose dynamics of interpersonal weights are described by equation (11). And the second type enclose the robot-agents, whose self-weights are static as well as their interpersonal weights. Therefore the relative interaction matrix $\mathcal{C}$ can be divided in four components:

$$\mathcal{C} = \begin{bmatrix} \mathcal{C}^{hh}_{n_h \times n_h} & \mathcal{C}^{rh}_{n_h \times n_r} \\ \\ \mathcal{C}^{hr}_{n_r \times n_h} & \mathcal{C}^{rr}_{n_r \times n_r} \end{bmatrix} \tag{12}$$

The meaning of each subscript and superscript of the relative interaction matrix $\mathcal{C}$ is given below:

- n_h: number of human-agents.
- n_r: number of robot-agents.
- hh: interpersonal relations human→human.
- hr: interpersonal relations human→robot.
- rh: interpersonal relations robot→human.
- rr: interpersonal relations robot→robot.

Thus, the total number of agents is given by:

$$n = n_h + n_r \tag{13}$$

The submatrices of $\mathcal{C}$ have their elements defined as

$$\mathcal{C}^{hh} = [c^{hh}_{ij}] \ ; \ \mathcal{C}^{hr} = [c^{hr}_{ij}] \ ; \ \mathcal{C}^{rh} = [c^{rh}_{ij}] \ ; \ \mathcal{C}^{rr} = [c^{rr}_{ij}]$$

where the indices i and j denotes the absolute position of the element in the whole matrix $\mathcal{C}$. Since matrix $\mathcal{C}$ is row-stochastic with zero diagonal, the relative interpersonal weights must obey the following constraints:

$$c^{hh}_{ii} = 0 \ ; \ c^{rr}_{ii} = 0 \tag{14}$$

$$\sum_{j=1}^{n_h} c^{hh}_{ij} + \sum_{j=n_h+1}^{n} c^{rh}_{ij} = 1 \tag{15}$$

$$\sum_{j=1}^{n_h} c^{hr}_{ij} + \sum_{j=n_h+1}^{n} c^{rr}_{ij} = 1 \tag{16}$$

$$\sum_{j=1}^{n_h} c^{hr}_{ij} \triangleq \frac{1}{n-1} \tag{17}$$

$$\sum_{j=n_h+1}^{n} c^{rr}_{ij} \triangleq \frac{n-2}{n-1} \tag{18}$$

Hence, with the insertion of agents with different weights dynamics than the stipulated by equation (11), we desire to alter the output of an opinion formation process that considers a finite set of possible opinions.

It is unrealistic to assume that groups of humans will act synchronously when arguing about a given issue. Instead of agents simultaneously updating their opinions, it is possible to assume that they interact in pairs. This approach is known as gossiping, which consists of two agents interacting at any step and, after that, one or both of their opinions can be changed (PROSKURNIKOV; TEMPO, 2018). Hence, the model for handling discrete opinions, presented in Sections 2.4 and 4.1, is applied here to carry the debate over each issue, using the issue-dependent weight matrix, $W(s)$, to define the probabilities of interaction between agents.

4.1.2 Simulations for Humans Self-esteem Evolution

In this section, we present experiments consisting on computational simulations. One of the objectives of these experiments are to analyze the effect over the humans' self-esteem caused by *bots* infiltrated in the group. This is done based on a model widely used in the literature. Later, we want to analyze if this effect on self-esteem impacts also the prevailing consensus information, when considering our proposed model for consensus over finite sets.

We present two meaningful scenarios here. The first scenario consists of ten agents with weight-dynamics according to the DeGroot-Friedkin model and an asynchronous asymmetric gossip model is chosen to the discussion of issues. In the second scenario, we maintain the gossip model, however, two types of agents are assumed, being eight as in the first scenario and two with static self-weights.

Then, we analyze the opinion consensus, the number of iterations, and the self-weights evolution. Furthermore, we consider whether the fact of some agents self-confidence being stationary can affect the output of discussed issues. Lastly, the results with a larger group are shown to demonstrate the model scalability.

Algorithm 1 represents the pseudocode to carry the simulations, separated into three stages: parameter settings (beginning), issues iterations (outer loop) and consensus iterations (inner loop). It is relevant to notice that the self-weights dynamics influences the opinion consensus, but the contrary is not true.

4.1.2.1 Discussion

The consensus outputs are observed in Table 4 for the two proposed scenarios described previously. The letters in bold for the initial opinions indicates which agents were replaced for robotic agents in the corresponding scenario. As can be verified the presence of robot-agents changed the consensus information in every case, except the last one. Another interesting trait that can be observed is that, in all issues, when the robots are considered the consensus output was the initial opinion from one of them.

Thus, in this case, changing the dynamics of only two agents in the system demonstrate that it is possible to alter the consensus opinion, even when that informa-

Algorithm 1 Steps of the computational experiment

```
Set the parameters of the experiment:
define number of humans, $n_h$, number of robots, $n_r$, and the initial self-weights of all nodes,
$\chi_1 = [\chi_1^1, \dots, \chi_1^{n_h}, \dots, \chi_1^n]$, where $n = n_h + n_r$

create a random row-stochastic matrix $\mathcal{C} = [c_{ij}]$ that respect the specifications given in ((14)-(18))

compute the eigenvector centrality scores, $\sigma \in \mathbb{R}^n$, for the matrix $\mathcal{C}$ as described previously
--------------------------------------------------------------------------------
Outer loop, issue iterations:
$s = 1$
while $|\chi_s - \chi_{s-1}| < \xi$ do
    set the initial opinions of all nodes, $y(s,1)$

    compute the influence matrix $\mathcal{W}_s = [w_{ij}(s)]$
    $w_{ii}(s) = \chi_s^i$
    $w_{ij}(s) = (1 - \chi_s^i) c_{ij}$

    compute a confidence range matrix $\mathcal{F}_s = [f_{ij}(s)]$
    $f_{ij}(s) = \sum_{q=1}^{j} w_{iq}(s), \; i,j = 1, \dots, n$
--------------------------------------------------------------------------------
    Inner loop, consensus iterations:
    $k = 1$
    while $\exists (i,j) : z_{s,k}^i \neq z_{s,k}^j$ do
        create a random normalized vector $\varphi \in \mathbb{R}^n$
        for $i = 1, \dots, n$ do
            for $j = 2, \dots, n$ do
                if $\varphi^i > f_{i(j-1)}(s)$ and $\varphi^i < f_{ij}(s)$ then
                    $x_{s,k+1}^i = x_{s,k}^j$
                else if $\varphi^i < f_{i1}(s)$ then
                    $x_{s,k+1}^i = x_{s,k}^1$
                end if
            end for
        end for
        $k = k + 1$
    end while
--------------------------------------------------------------------------------
    the human nodes update its self-weights by the DeGroot-Friedkin model ((11))

    the robotic nodes persist its self-weights across issues
    $\chi_{s+1}^i = \chi_s^i, \; i = n_h + 1, \dots, n$

    $s = s + 1$
end while
```

Table 4 – Consensus outputs.

Issue $[s]$	Initial Opinions $[x_{s,1}]$	Final Opinions without robots	Final Opinions with robots
1	{B,A,A,C,A,C,B,C,**A**,**B**}	C	A
2	{B,C,C,B,B,B,A,B,**A**,**C**}	B	C
3	{A,C,A,A,B,A,C,B,**A**,**A**}	C	A
4	{B,B,C,B,A,C,A,B,**B**,**C**}	C	B
5	{A,A,B,B,A,B,C,C,**C**,**B**}	B	C
6	{B,C,C,A,A,A,A,C,**A**,**C**}	A	A

tion is not prevalent in the beginning. Furthermore, Figure 35 shows that, from issue six onward, the self-weights stabilize each one at a different state. Hence, an equilibrium is reached but with distinct values for each agent. Although the robot-agents have static self-weights, it is important to emphasize that their opinion also changes according to the gossip model, as occurs with the humans.

The case when all agents' self-weights converge to the same value $(1/n)$ is called a democratic configuration (JIA et al., 2015) and it requires the relative interaction matrix C to be doubly stochastic.

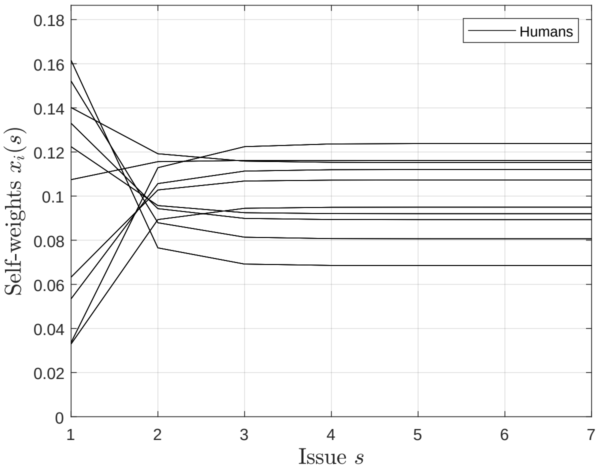


Figure 35 – Self-weights evolution with 10 agents (humans).

In the second scenario, Figure 36, it is observed that the self-weights stabilize each one at a different state from issue four onward. However, this time, all human-agents have a considerably lower self-weight when compared to the first scenario. In other words, the presence of agents with static self-confidence resulted in the other agents self-depreciation.

The connections of all agents and their respective self-weights evolution between issues can be seen in Figure 37, in which the edge's width indicates the interpersonal weight, and the node's size represents the self-weight value. The first graph presents the initial conditions, and the second one depicts the states when the self-weights dynamics have already stabilized.

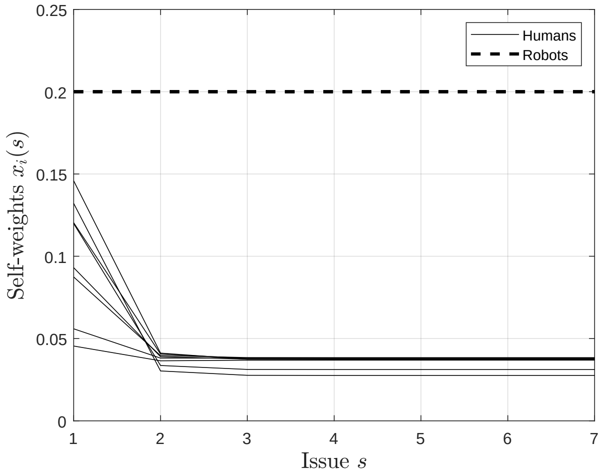
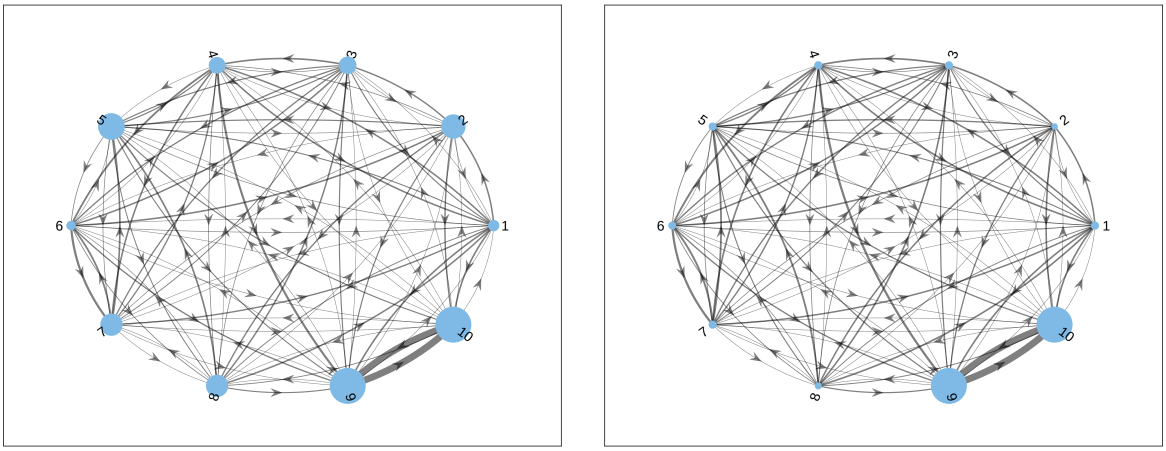


Figure 36 – Self-weights evolution with 8 human agents and 2 robot agents.



(a) initial conditions at $s = 1$

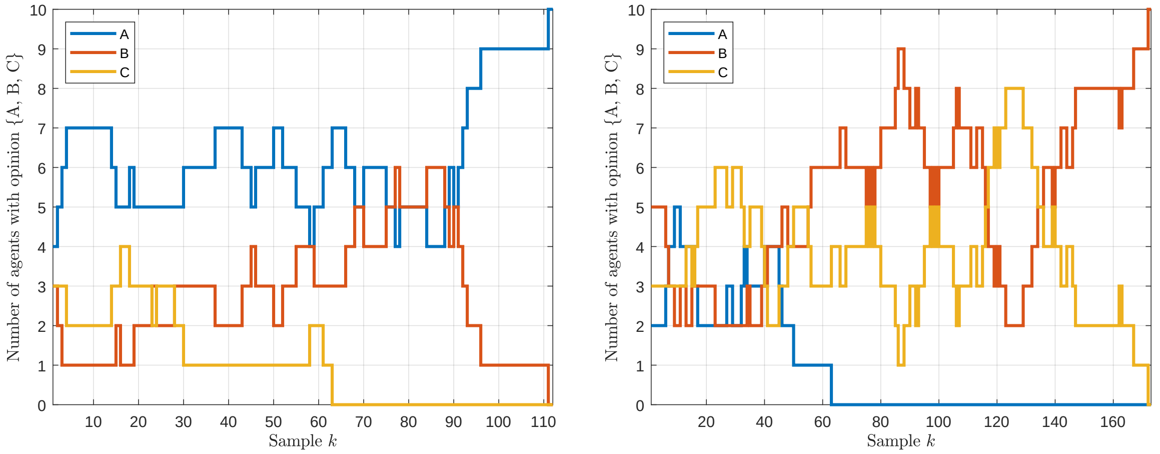
(b) static state conditions at $s = 4$

Figure 37 – Evolution of the influence network across the sequence of issues.

Following the opinions evolution, Figures 38a and 38b show how many nodes share the same opinion and how many steps took to reach consensus for the initial self-weights and after their stabilization, respectively. It is important to emphasize that the two dynamics (consensus and self-weights) do not interfere in each other and it is possible to analyze them separately.

It can be seen that, for the weight's initial conditions, it was almost twice faster than when compared with the number of steps necessary to reach consensus when the self-weights have already stabilized. One hypothesis for justifying this phenomena is that less self-confident human-agents will more easily change their opinion, what may delay the consensus.

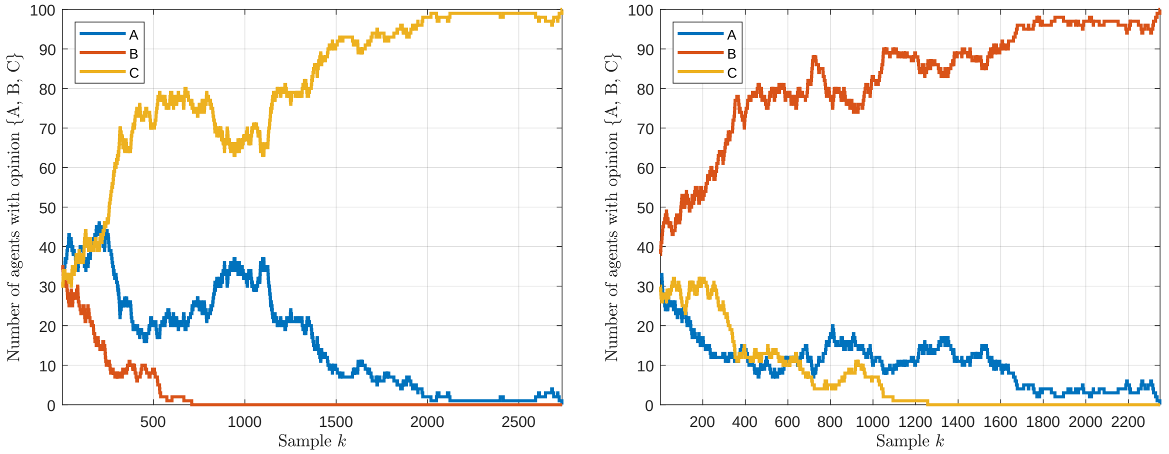
Finally, to demonstrate the scalability of the proposed analyses, we present a scenario with one hundred agents, such that 90 of them are humans and 10 are robots. The first issue discussion is shown in Figure 39a. Figure 39b depicts the discussion in the last issue.



(a) Initial self-weights ($s = 1$)

(b) Stabilized self-weights ($s = 4$)

Figure 38 – Opinion evolution with 8 human agents and 2 robot agents.



(a) Initial self-weights ($s = 1$)

(b) Stabilized self-weights ($s = 3$)

Figure 39 – Opinion evolution with 90 human agents and 10 robot agents.

When compared to the first two simulations, this third case reaches consensus much more slower, both for the self-weight's initial and final states. This is in agreement to the notion that small groups will come to a consensus faster than bigger groups. However, the main feature here is that, even with one hundred agents, the system reaches an opinion agreement, meaning that the model is scalable to handling swarms of robots interacting with huge social networks.

The analyses presented here demonstrate the possibility to change the consensus opinion by just influencing indirectly the self-confidence of the agents. Once this is verified, there are many open questions to be explored in future works. As examples, it is possible to investigate on how to define the best dynamics for the robotic agents and whether it is possible, by manipulating these dynamics, to guarantee the convergence for a desired consensus information.

Also, it may be studied the possibility of delaying or accelerating the consensus by manipulating the interaction weights accorded by robotic agents to the humans' opinions. Furthermore, the influence of robotic agents' placement when using different types of topology could be explored. Additionally to all these propositions, extensions to the DeGroot-Friedkin self-appraisal model could be prospected.

One final and interesting proposal for further works is to develop a real-world experiment, where a group of volunteers would have to discuss a few issues through some interface and come to an agreement, without knowing that there are also robotic agents participating in the debate. Despite the great ambition of this experiment, it could bring key results for understanding how contemporary society can be misguided and, hopefully, give some insights on how to avoid it.

4.1.3 Taking Reliability into Account

In order to add some notion of reliability to the information, it will be considered as a 2-tuple

$$X_k^i = \langle m, \tau_m \rangle$$

where m is the content of the information from agent i; and τ_m is a numeric measure of the reliability of this information.

Next, a sample dependent matrix is defined to couple the influence of one agent over others with the reliability of the information held by the agent, such that,

$$\mathcal{A}_k = [\alpha_{ij}(k)], \;\; \alpha_{ij}(k) = w_{ij} X_k^{j\langle 2\rangle} \;, \;\forall i{,}j$$

where w_{ij} are the influence weights described before and $X_k^{j\langle 2\rangle}$ is the second element of the tuple, i.e., the reliability coefficient of the information within agent j at the sample k.

Since the reliability coefficients, τ_m, multiply the influence weights, ranges can be defined for them, such that

$$\tau_m \in \mathbb{R}_+ \rightarrow \begin{cases} 0 \leq \tau_m < 1 & \text{, information is not reliable} \\ \tau_m = 1 & \text{, the reliability is irrelevant} \\ \tau_m > 1 & \text{, information is reliable} \end{cases}$$

The matrix $\mathcal{A}(k)$ is not row-stochastic and, in order to be used as the matrix of probabilities of interaction, it needs to be normalized as follows

$$\mathcal{P}_k = [p_{ij}(k)], \text{where } p_{ij}(k) \triangleq \frac{\alpha_{ij}(k)}{\sum_{j=1}^{n} \alpha_{ij}(k)}$$

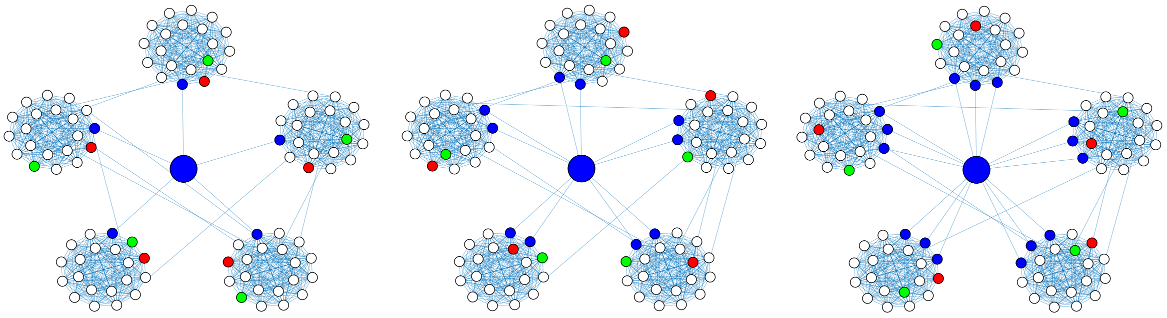
4.1.4 Simulations for Evaluating the use of a Reliability Stamp

The experiment shown here also consists of computational simulations but aiming to investigate the influence of malicious nodes in the spreading of a desired information over the network. Also, we propose the implementation of a stamp to mark if specific information is reliable, and the effects of this policy are analyzed.

The network configurations considered for those simulations were grounded in five fully-connected groups of twenty agents and an isolated node that only transmit information for some specific nodes in the groups. This isolated node could represent a mass media vehicle or a newspaper. However, for the present simulation, it will be considered as a central generator of malicious content, and the agents linked to it are robots infiltrated in the groups, accountable for sharing the information from the central node.

Three network configurations will be observed, differing from one another by the number of malicious agents per group. Also, in each configuration, there are ten random connections between non-malicious agents from different groups.

There are three types of information flowing through the network, represented by the colors blue, green, and red. When a node had not received any information yet, it will be represented by the color white. Figure 40 shows the initial states for each topology, and it can be noticed that each group starts with only one node holding each of the non-malicious information plus the malicious nodes.



(a) One malicious node per group (b) Two malicious nodes per group (c) Three malicious nodes per group

Figure 40 – Initial conditions for each scenario.

Hence, at the first moment, we carried 1000 simulations for each network configuration, aiming to reduce the effect of randomness from the results and obtain more meaningful conclusions from it. The information will be exchanged between agents following the model from the previous section, and considering that all information had the same reliability coefficient, i.e., $\tau_m = 1 \ \forall m$.

Furthermore, we imposed a limit of three changes of information for each agent, modeling that once it made its mind about an issue, it will not change anymore. The results collected from this first set of simulations are summarized in Table 5; it is worth mentioning that the averages were rounded to the nearest integer.

From these results, it is noticeable that the spreading of the malicious information increases proportionally to the number of malicious agents per group.

Later, we considered a reliability stamp for the information, meaning that a certain information can be marked as reliable. If an information is marked, then the reliability

Table 5 – Results averaged over 1000 simulations for each scenario, considering the blue information as malicious and no information with reliability stamp.

Number of Malicious Nodes	Blue Information (Bots)	Green Information	Red Information
1	48	26	26
2	69	16	16
3	80	11	10

coefficient associated is set to $\tau_m = 2$.

Once more, we carried 1000 simulations for each scenario, but marking the information represented by the color red as reliable. Namely, the reliability coefficients were set as $\tau_{blue} = \tau_{green} = 1$ and $\tau_{red} = 2$. In Table 6, the result of this second set of simulations is summarized.

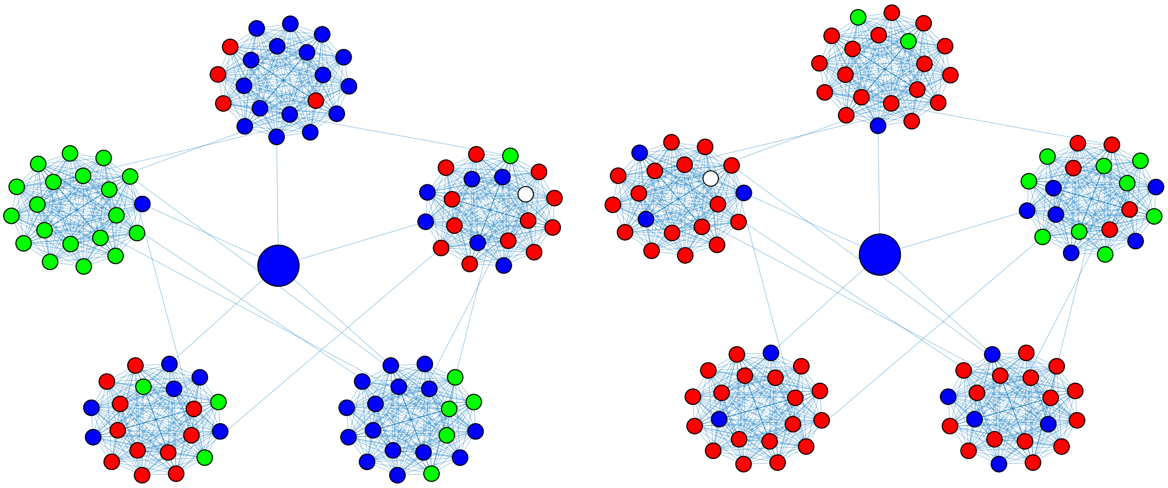
Table 6 – Results averaged over 1000 simulations for each scenario, considering the blue information as malicious and the red information as marked by the reliability stamp.

Number of Malicious Nodes	Blue Information (Bots)	Green Information	Red Information (Marked)
1	21	9	70
2	38	7	55
3	52	6	42

The effect noticed was that the reliability stamp is able to reduce the spread of malicious information in the network. Still, the number of malicious agents per group influences the spreading. Another remark about these results is that for more than two malicious agents per group, the spreading of the malicious information surpasses that of the marked information.

It may be necessary to emphasize the fact that the bots infiltrated in the groups do not update their information with information from other members of the group. Hence, the malicious nodes are stubborn in the blue information throughout the simulations.

Figures 41 to 44 illustrate the difference made by the applying the reliability stamp; they picture one of the simulations carried in each scenario. It is important to emphasize that the graphs presented in Figures 41 to 44 do not match the result from Tables 5 and 6 since these are averaged over all the simulations carried. Also, in Figure 42, we bring a QR Code to access an animation for the spreading with one bot per group and no reliability stamp.

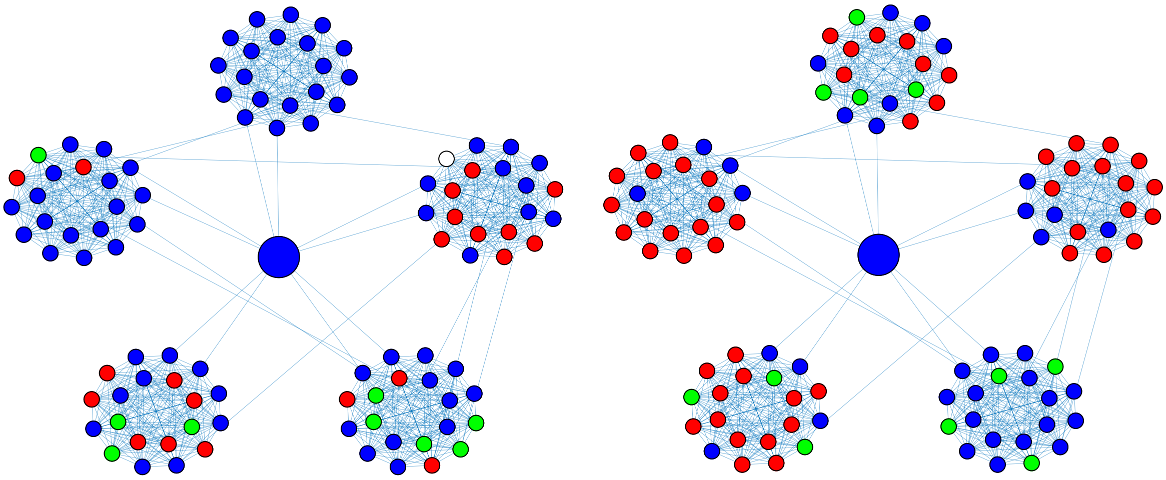


(a) Without reliability stamp

(b) With reliability stamp

Figure 41 – Example of final states in the scenario with one malicious node per group.

Figure 42 – QR Code directing for the animation exemplifying the simulation in the scenario depicted in Figure 41a, also accessible through the link: https://youtu.be/0XXX9Q_jeNk.



(a) Without reliability stamp

(b) With reliability stamp

Figure 43 – Example of final states in the scenario with two malicious node per group.

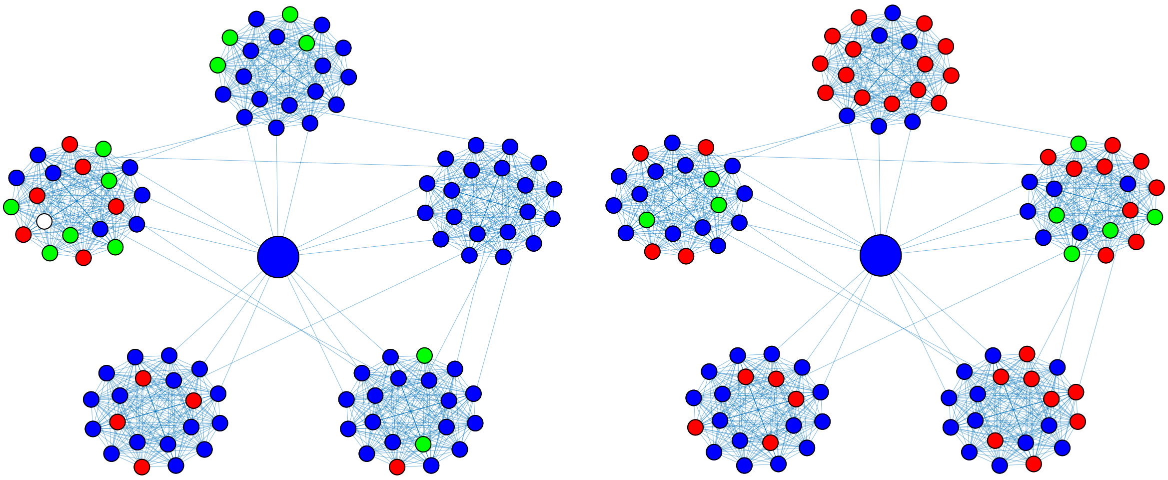

(a) Without reliability stamp (b) With reliability stamp

Figure 44 – Example of final states in the scenario with three malicious node per group.

The implementation of control measures and utilization policies are important to safeguard ethical principles in the use of OSNs. Here, we applied dynamic models that represent the diffusion of information and opinions in networks to evaluate the influence of *bots* in the opinion evolution when compared with reliable information.

In the simulated examples, it was verified that the action of bots decisively influences the information diffusion and the opinion consolidation in the analyzed networks. We also conclude that the adoption of a mechanism to attribute some degree of reliability to the information sent to the network could counteract the influence of malicious nodes (*bots*). However, this impact could be neutralized with the increase in the numbers of bots.

Finally, by the results obtained in this work, adding measures to detect false news and implement information certification will only be effective by jointly implementing policies to detect and exclude bots from instant messaging application networks.

4.2 ANALYZING THE CENTRALITY OF THE NODES

The condition to ensure that the model from Sections 2.4 and 4.1 reaches consensus is given by the eigenvalues of $\mathcal{W}$. More precisely, it is dependent on the algebraic multiplicity of its largest eigenvalue. Let $\Lambda = [\lambda_1 \ldots \lambda_n]^\top \in \mathbb{C}^n$ be a vector containing the ascending ordered eigenvalues of $\mathcal{W}$. Then, (5) converges if and only if $\lambda_n = 1$ is unique in modulus, i.e., the following inequality holds

$$|\lambda_1| \leq |\lambda_2| \leq \ldots \leq |\lambda_{n-1}| < \lambda_n = 1.$$

Even though we can check that the model converges, we can not predict the final value of the consensus as in the classical deterministic model. However, we may use some knowledge about the network structure to predict the probabilities of converging to a given piece of information. It is impossible to visually gather information from large and complex networks, we need to rely on metrics such as node centrality, average path length, and clustering coefficient.

In that sense, we state that the eigenvector centrality dictates the probability of a given information to prevail in the consensus performed by our model. Further, this indicates that the more central in the network one person is, the higher is hers/his capacity of spreading an information. The experiments presented in this chapter corroborate with these claims. Also, we refer the reader to the PhD thesis by Carvalho (2020), our colleague from the research group, in which this aspects are further explored.

4.2.1 Eigenvector Centrality

There are many centrality measures such as Katz centrality (KATZ, 1953), and the Page Rank centrality (GLEICH, 2015), to say some. This work is mainly interested in the eigenvector centrality because of its relation with some network algebraic properties, which will be shown in simulations and experiments.

The eigenvector centrality score is a metric that quantifies the influence of a node over the network, in other words, it measures how many connections a node has and how much its neighbors appreciate its opinion. This measurement was first proposed by (BONACICH, 1972), and since then, it is widely adopted to define the relative importance of an individual in a social network.

Consider a graph $\mathcal{G}$ with n nodes and its normalized adjacency matrix $\mathcal{W} \in \mathbb{R}^{n \times n}$. A left eigenvector corresponding to the eigenvalue λ_i is a vector $v_i \in \mathbb{C}^n$ satisfying $v_i^\top W = \lambda_i v_i^\top$.

The eigenvector centrality scores are calculated as

$$\sigma = v_{max} / \|v_{max}\|_1 \tag{19}$$

where $v_{max} \in \mathbb{R}^n$ is the dominant left eigenvector, i.e., associated with the maximum eigenvalue (for a row-stochastic matrix, $|\lambda_n| = 1$). The i-th element of $\sigma = [\sigma^1 \dots \sigma^n]^\top$ is called the centrality score of agent i.

4.2.2 Simulations for Estimating the Probability of Prevalence

In this section, we perform a large number of computational simulations to test if the model proposed in Sections 2.4 and 4.1 is capable to reach consensus in the same scenarios that others found in the literature are. These simulations also explore situations where other algorithms are not able to succeed. The simulation is over when the group reaches consensus in one of the initial opinions. We call consensus' opinion the one that prevailed at the end of the simulation. The results corroborate our proposed hypothesis that the eigenvector centrality (BONACICH, 1972) of the network indicates the probabilities of each initial opinion becoming the consensus' opinion for the group.

In order to analyze if the algorithm works under real-world adversities and limitations, experiments were done using a group of five robots acting as communication nodes. It was chosen a topology with one node that have its centrality score hugely discrepant to help in identifying a tendency, even over a small number of iterations.

To carry the simulations, it is also needed to define a matrix, $\mathcal{F} = [f_{ij}] \in \mathbb{R}^{n\times n}$,containing the interaction ranges of each agent. These ranges are useful in the computational simulations to randomly select the interactions obeying the distribution defined by $\mathcal{P}$ (interaction probability matrix, see (6)). It is used for the specific way in which the algorithm were designed. Hence, each element of $\mathcal{F}$ is computed as

$$f_{ij} = \sum_{q=1}^{j} p_{iq},\ i,j = 1,\ldots,n \tag{20}$$

4.2.2.1 Computational Simulations

Here, the results of numerical simulations to support our observations presented previously are shown. The simulations are executed on different networks to evaluate the effect of the topology on opinion diffusion, and how it influences the eigenvector centrality and the spreading probability for each opinion. Algorithm 2 represents the pseudo-code to carry out the simulations.

Twelve different topologies were chosen for testing the algorithm. In each scenario, 100,000 simulations were run to get some statistic knowledge about the algorithm. The initial opinions were the same in all simulations using the same topology. Each agent started with a unique letter from the alphabet as its opinion. The algorithm led the group to a consensus in every simulation run.

The scenarios used can be observed in Figure 45, where the nodes are presented, with their initial opinion, and the connections between them. The appraisal that the agents have to its neighbors, in other words, the values of each connection respect the restriction of the row-stochastic matrix. The components of this weighted adjacency matrix are the probabilities of interaction between the agents. This means each agent assigns equal weights for the opinion of every neighbor, including their own.

Some of the topologies presented in the previous topic can be highlighted here, as some of them have different behavior when it is considered the classic gossip algorithms or even the DeGroot model. As can be verified, topologies (b), (e) and (l) are unable to achieve consensus using these models.

This occurs because the weighted adjacency matrix has its largest eigenvalue with algebraic multiplicity greater than one. The behaviour observed in these scenarios are the interchange of two remaining opinions in the network in an oscillatory way.

Algorithm 2 Steps of the computational experiment

```
Set the parameters of the experiment:
define number of agents $n$ and number of simulations $n_s$.

create the row-stochastic matrix $\mathcal{W} = [w_{ij}]$ from an arbitrary unweighted adjacency
matrix $\mathcal{A}$. The computation of each element of $\mathcal{W}$ is done by normalizing each element of $\mathcal{A}$.

compute the confidence range matrix $\mathcal{F} = [f_{ij}]$ as in (20)

Simulation iterations:
for $s = 1, \ldots, n_s$ do
    Consensus iterations:
    $k = 1$
    while $\exists(i,j) : x_k^i \neq x_k^j$ do
        randomly select the active agent $i_k \in \{1, \ldots, n\}$
        following an uniform distribution
        $i = i_k$
        generate a single uniformly distributed random
        number, $\varphi$, in the interval $(0,1)$ for picking the
        agent $j$ for interacting with $i$ in the following loop
        for $j = 2, \ldots, n$ do
            if $\varphi > f_{i(j-1)}$ and $\varphi < f_{ij}$ then
                $x_{k+1}^i = x_k^j$
                break
            else if $\varphi < f_{i1}$ then
                $x_{k+1}^i = x_k^1$
                break
            end if
        end for
        $k = k + 1$
    end while
    Store the occurrence of each consensus information in
    vector $\vartheta \in \mathbb{N}^n$
end for

compute the eigenvector centrality scores, $\sigma \in \mathbb{R}^n$, for matrix $\mathcal{W}$ and compare to the
occurrence ratios, $\vartheta / n_s$.
```

It is relevant to say that this happens even when the algorithms are averaging over real-valued opinions.

The algorithm presented here, as stated before, is able to reach consensus in those cases. This happens due to asynchrony, a feature usually undesired but that enables the network to break out of those oscillatory states.

As claimed previously, we desire to test if the centrality of a given agent can be

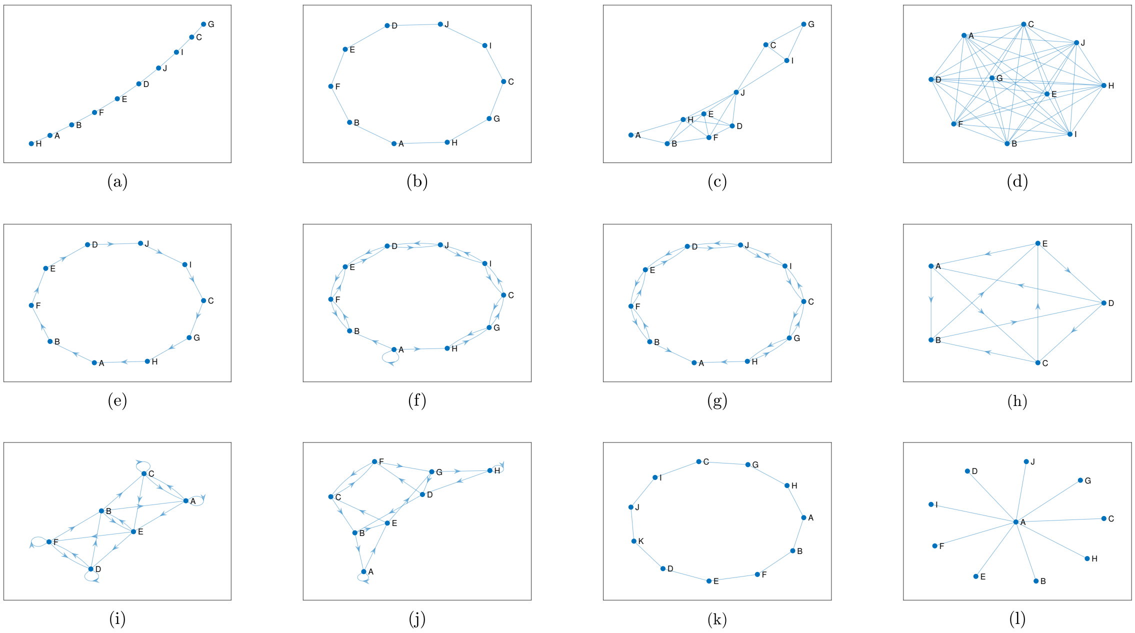

(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l)

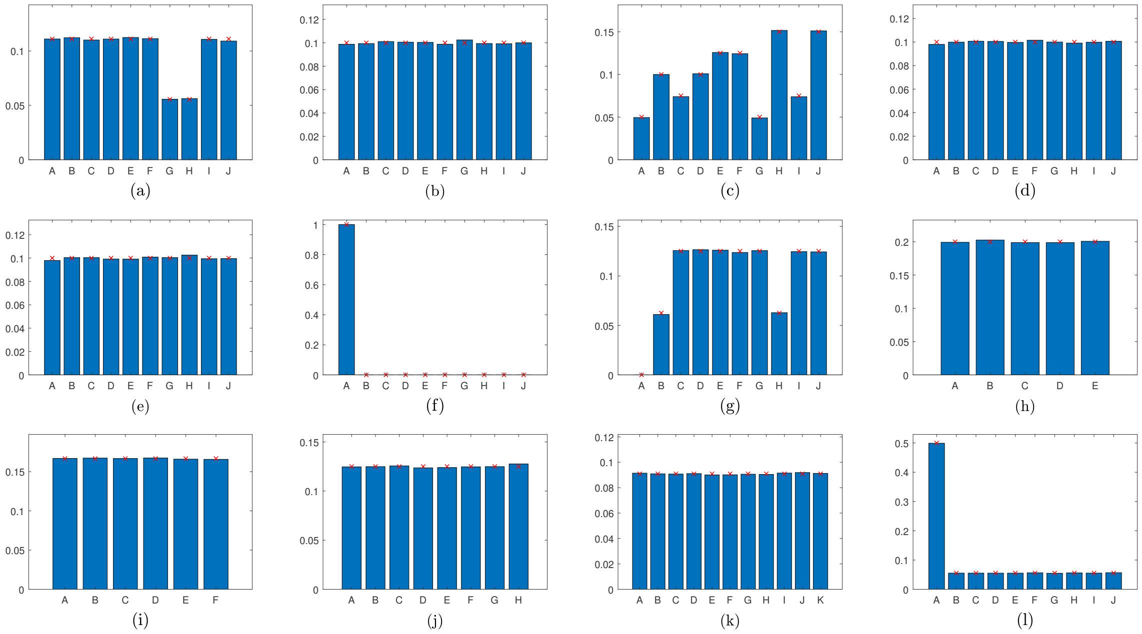
Figure 45 – Topologies: (a) Linear graph (b) Circular graph with even number of nodes (c) Arbitrary graph (d) Full connected graph (e) Circular digraph with clockwise direction (f) Circular digraph with source (g) Circular digraph with sink (h) Five nodes, each with two connections (i) Six nodes, each with three connections (j) Eight nodes, each with two connections (k) Circular digraph with odd number of nodes (l) Star graph.

a way of predicting the probability of its initial opinion becoming the consensus opinion for the group. In the carried simulations it was also computed an occurrence ratio, that is, how many times each letter was chosen as the consensus opinion over the total number of simulations.

It is possible to observe in Figure 46 that the occurrence ratio is very close to the eigenvector centrality scores. The error between these two values goes to zero as the number of simulations increases. This is an important result for the proposed algorithm, giving some sense of predictability in its behavior. Following, it will be presented a more detailed analysis of the simulations results.

First, consider the linear topology (Figure 45 (a)) in which all nodes have the same number of neighbors with the exception of the both ends. The nodes with two neighbors have centrality score $\sigma^i = 0.1125$ and the nodes with just one neighbor have $\sigma^i = 0.05$. As expected the percentage of occurrences were very close to their respective centrality scores.

The topologies showed at Figure 45 (b), (d), (e), (h), (i), (j), (k) have the same centrality scores for all nodes in their topology $[0.100, 0.100, 0.100, 0.200, 0.166, 0.125, 0.090]$, respectively. Thus, for these systems all nodes have the same amount of importance in the network, this way the $\sigma = \mathbf{1}_n/n$, where n is the number of nodes in the network. Again all percentage of occurrences were near the expected value.



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l)

Figure 46 – Comparison between the occurrence ratio of each initial opinion and the centrality score (red crosses) of the agents in the respective topologies from Figure 45.

The topology (c) is the one with more variability in the centrality scores, in which nodes H and J have the greatest influence and the nodes A and G are the least influential. One more time the probabilities approached the centrality scores.

The topologies (f) and (g) have a source and a sink respectively. The sink have $\sigma^i = 0$ and it is not possible to reach consensus in this opinion. Also, the neighbors of the sink node will have lower influence than the others, since the sink node has no influence in the network. And the case with a source node is quite the opposite with the source node with $\sigma^i = 1$ and all the others with $\sigma^i = 0$, indicating that the consensus will always converge for the opinion of the source node.

Finally the topology (l) is the example in which one node has much more influence than all the others, as the central node communicates with everybody and all the others communicates only with the central node. Thus in this case the σ^i for the most influential node is equal to 0.5 and for all the others is equal to 0.1. As expected the probabilities were nearly the centrality scores.

Therefore, the simulations corroborated with the hypothesis that the eigenvector centrality scores can be interpreted as the probability of a certain opinion to occur in the end of the consensus process.

4.2.2.2 Laboratory Experiments

In this section, we perform experiments intending to validate our numerical observations in a practical scenario, where there are communication issues and other physical limitations that may interfere in the information diffusion. The experiments were performed with Kilobots, a low-cost robot platform designed originally by researchers at Harvard University to make the test of algorithms for collective robots accessible to researchers worldwide (RUBENSTEIN; AHLER; NAGPAL, 2012).

The Kilobots used in the experiments are a personalized version, developed by professors and students from the Department of Automation and Systems at the Federal University of Santa Catarina[1]. The experiments involved five robots arranged in a star topology with self-loops, as described by Figure 47. All border robots can communicate only with the central node, which can reach everyone else.

The robots are all static in a fixed position (they do not move) which allows them to have the communication topology as described by Figure 47b. Under this configuration, the eigenvector centrality becomes $\sigma = [0.385\ 0.154\ 0.154\ 0.154\ 0.154]$, indicating node 1 as the more influential over the network and the other nodes as equally influential

The communication issues change the consensus probability for each initial opinion, which differs from the eigenvector centrality score, since, in our case, the transition matrix does not take into account those issues. However, the expected communication issues are limited and relatively small: according to Rubenstein, Ahler, and Nagpal (2012), in an experiment with 25 robots, the communication channel could support on average 32% of usage for five-byte packages without packet loss due to collision. In our case, there are only five robots, and their communication packages have 9 bytes. Despite the influence of communication issues, we expect that the results will not deviate significantly from those obtained in the simulations.

Each robot has its opinion at the beginning of the experiment set as $z_0 = [\text{A B C D E}]$, i.e., all robots have distinct opinions. Such opinions are represented through five different colors that are displayed by the robots through a led: A is purple, B is red, C is green, D is blue, and E is black (light off).

To verify our hypothesis, we perform three tests[2] composed by sets of observations looking for the prevailing opinion after the consensus be reached. Each test has exactly 50 executions of the asynchronous gossip algorithm under the same initial conditions. By the law of large numbers, we expect that as the number of execution increases, the error between the expected probability of consensus over a particular opinion and the fraction of occurrence of each opinion decreases satisfactorily.

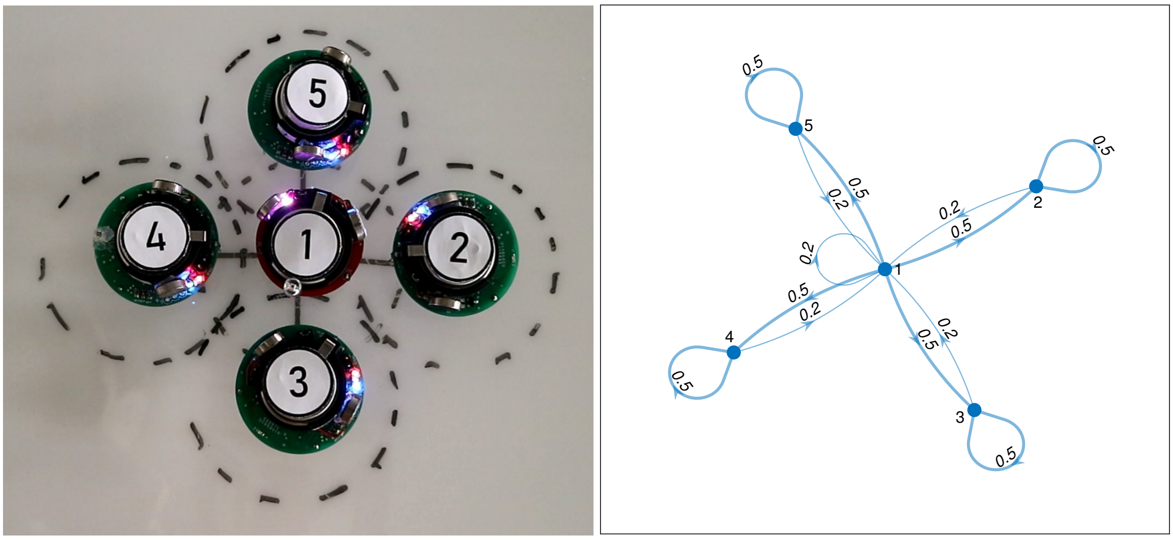
Table 7 contains the occurrence of each opinion at the consensus and the expected values given by the eigenvector centrality. As one can see, consensus in opinion A had a greater occurrence ratio than the other opinions for all tests.

By the prediction of eigenvector centrality, opinions B, C, D, and E should have

---

[1] `http://kilobots.paginas.ufsc.br/`

[2] Videos available at `https://youtube.com/playlist?list=PLJFozF36QU0LdclyP7nlJt9bhr2-lq0qX`



(a) Top view of the robots (b) Underlying topology graph

Figure 47 – Adopted topology. Dashed circles represent an estimation of Kilobots communication range, continuous lines are logical connections between them, and the edge weights represent the importance each robot attributes to its neighbors.

Table 7 – Results from the experiments with kilobots.

Opinion	**Predicted**	**Test 1**	**Test 2**	**Test 3**
A (purple)	38.46%	38%	34%	48%
B (red)	15.38%	22%	22%	20%
C (green)	15.38%	14%	12%	10%
D (blue)	15.38%	18%	18%	10%
E (black)	15.38%	8%	14%	12%

the same probability of occurrence, since the robots that start with these states have the same importance to robot 1 (the center of the star). There is an error between the prediction and occurrence for each opinion: the most significant errors happen at opinion E (7.38%) in Test 1, opinion C (3.38%) in Test 2, and opinions C and D (5.38%) in Test 3. Note that, the increase in the prevalence of opinion B may indicate that link from robot 1 to robot 2 was weaker (due to packet loss or other physical factors) than other links in the network. The same interpretation can be used to explain why in Test 3, opinion A was more frequent than what it is supposed to be, which is probably the result of communication issues.

There may be some variations from one test to another, e.g., lower battery levels. However, the three tests are essentially sessions of the same experiment, allowing us to analyze their outcomes together. Hence, considering the 150 executions of the asynchronous gossip algorithm, we get an aggregate occurrence ratio for each information. In Figure 48, we can see the absolute errors of these outcomes in relation to the values predicted by the eigenvector centrality.

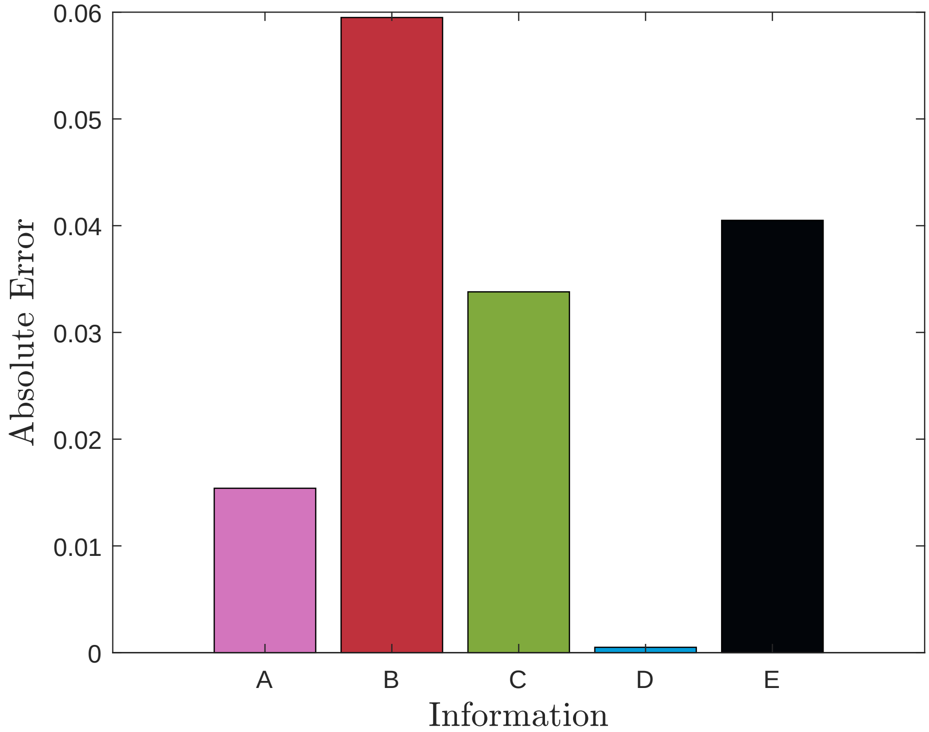


Figure 48 – Absolute errors between the eigenvector centrality scores and the aggregate outcomes from all three tests.

Although these errors, the prediction is pretty accurate, given the number of simulations. If it were possible to increase substantially such value, and include in the probability transition matrix the fault probability for each link, it would be expected lower errors in the observations and a more precise validation of our hypothesis.

After all, the simulations and the experiments with Kilobots confirm the hypothesis that eigenvector centrality is indicative of the occurrence probability of each node's initial opinion when the group reaches consensus. For the best of our knowledge, it is the very first time such relation is pointed out in the literature.

We show, through experiments, evidences the proposed algorithm can achieve consensus over the same conditions that previous works did. Also, it can handle discrete opinions and reach consensus in scenarios where synchronous approaches present oscillatory responses. At last, it was confirmed that there is a relation between the eigenvector centrality scores and the probability of reaching consensus on a particular opinion.

For future works, we intend to achieve formal mathematical proofs for the convergence of the algorithm and the relationship between algebraic graph theory properties and the probabilities to reach consensus on a certain opinion.

Also, as mentioned in the paper, the presented algorithm has the weights of the connections as the only impacting factor in the final consensus opinion for the network. Hence, there is room for investigation in ways of considering the content of the opinions and not only the influence that agents have in each other. For applications in robotics, factors like the precision of sensors from each robot may be embedded in the weighted adjacency matrix.

5 CONTROLLING THE SPREAD

Online personalization is a process that comprehend many aspects of individualizing the interaction and content exchanged by a system and its users (ZANKER; ROOK; JANNACH, 2019). One representative area within the personalization research is the study of Recommender Systems (RSs). According to Ricci, Rokach, and Shapira (2015), these systems are software capable of suggesting items that are more related to the interests of a specific user. To list a few examples of these suggestions, they could be what products to buy, what music to listen, what movie to watch, what news to read, what social-media content to interact with, etc.

Specially, when we speak about the recommendation of news articles and social-media content, we are dealing with a process that can shape the way people see the world (PARISER, 2011). However, this is not necessarily profitable for the companies running these RSs. Their income depends highly on how much (personalized) advertises they are able to show the users. This, in turn, is proportional to the amount of screen time users devote to the app or website owned by these companies. Therefore, it is valid to assume that one of the main goals of a company in this business is to maximize the screen time of its users in order to increase its profits. And this can be achieved by directing content that stimulates feelings like the comfort of having your biases confirmed or even the anger caused by some sensationalist news about someone in the opposite political camp.

One way to achieve this kind of algorithmic personalization is through clustering both of content and users, determining which groups of people should receive certain kinds of information. Considering the users of an Online Social Network (OSN), clustering them is called community detection or identification in the specialized literature. We refer the reader to two surveys on that subject: the work by Fortunato (2010), which brings an extensive review over the methods and algorithms known by the time; and the recent paper by Baltsou, Christopoulos, and Tsichlas (2022), focusing on local community detection to address the scenarios where the networks structure is not fully known.

For the moment, we will focus on global community identification, i.e., partitioning the entire network into a given number of groups of nodes. Also, we consider valid to assume that communities may overlap in an application such as social networks. In fact, this has emerged during experiments with hard clustering techniques over networks with complex structures, when some nodes were classified as belonging to a cluster even though they did not have any neighbors in that same cluster. These results will be further discussed in Section 5.3.

In this scope, we highlight the work by Collingsworth and Menezes (2012), in which an algorithm is proposed for identifying communities based on local calculations of node entropy. The idea is that the entropy represents how much a node is “satisfied” with its current community. Their method also handles the case when one (or more) node is “uncertain” about the community it belongs to by assuming that it belongs to all the communities (overlap).

Another work worth mentioning proposes an algorithm to identify overlapping communities by maximizing a modularity function (i.e., a metric for the clusters’ quality), a mapping of the nodes into the Euclidean space, and an Fuzzy c-means (FCM) clustering technique, which outputs a membership degree from every node to each identified cluster (ZHANG, S.; WANG, R.-S.; ZHANG, X.-S., 2007). According to the authors, this modularity’s maximization problem can be reformulated as an eigenvector problem, which is akin to those solved in spectral clustering algorithms. Hence, the approach used by the authors of that paper is similar to the one we will present in the following sections. However, they consider nodes with high membership degrees to two or more communities as the overlap between these communities, and do not devise a solution for determining a hard clustering from these results.

In our proposal, we suggest a modification of the spectral clustering which uses Fuzzy c-means to perform a soft partitioning of the network’s nodes. After that step, we take the highest membership degree to assign every node to a single cluster. This allow us to perform a hard clustering while still having information about nodes in the “overlapping zone”.

However, within the networks’ structures present in our studies on Online Social Networks, we have faced some undesired outputs of the community identification algorithm. Namely, some nodes were assigned to a cluster in which they did not have any connections with other members of that community. This led us to develop a correction step in our algorithm that uses the membership degrees and two-hops neighborhood information to re-assign these nodes (or pairs of nodes) that were isolated from their clusters.

Also, we envision that the knowledge gathered by applying our community detection algorithm can be used to develop policies for counteracting the spread of misinformation and/or promote the true information regarding some deceptive content flowing through the network. In Section 5.3, we present the results of using this knowledge about the identified communities together with a node centrality metric to determine the fastest way of organically spreading a piece of information through the network. These results differ from those presented in Chapter 4 because it surpass the analysis’ field and gives a first step in the direction of synthesis, since it can be used to suggest amendments to the network’s structure.

The contributions presented in this chapter can be summarized as follows: the development of an algorithm for performing community identification over complex networks; the analysis of clustering quality in comparison to other algorithms; the proposal of policies for controlling the spread of information based on the output of that algorithm; and a comparative study on the spreading capability between the different policies presented.

5.1 BACKGROUNDS ON CLUSTERING

We present here the algorithms used for building the technique we propose in Section 5.2 for community identification in social networks. It is a brief summary to allow fully understanding of our proposal.

5.1.1 K-means Clustering

Considering the problem of identifying groups of data points in a multidimensional space, the K-means algorithm is a procedure that solves it by minimizing an objective function called *distortion measure*. It is a two-stage optimization that alternates between finding the data points closer to each cluster's center and finding new clusters' centers over the previously computed clustering. This two-stage optimization repeats until convergence (BISHOP; NASRABADI, 2006).

Let $\{\theta_1, \ldots, \theta_N\}$ be a data set of N observations of a D-dimensional variable θ. The goal of this algorithm is to partition the data set into K clusters, assuming that K is given. First, we introduce a set of D-dimensional vectors μ_k, with $k = 1, \ldots, K$, such that μ_k represents the center of the k^{th} cluster. Then, the goal can be rephrased as finding an assignment of data points to clusters, and also a set of centroids $\{\mu_k\}$, such that the sum of the squares of the distances of each data point to its closest vector μ_k is a minimum.

Each data point θ_n is associated with a set of binary variables $r_{nk} \in \{0,1\}$, $k = 1, \ldots, K$ indicating in which of the K clusters θ_n falls in. Hence, if data point θ_n is assigned to cluster k then $r_{nk} = 1$, and $r_{nj} = 0, \forall j \neq k$. Therefore, we can write the objective function as

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|\theta_n - \mu_k\|^2 \tag{21}$$

which can be called a distortion measure, as we stated previously. So, the goal is to find values for $\{r_{nk}\}$ and $\{\mu_k\}$ that minimize J.

We can achieve that goal by iteratively performing two successive optimizations with respect to the r_{nk} and the μ_k. In the first iteration, we can choose values for the μ_k (e.g., k random elements from the data set), fix these values, and minimize J with respect to the r_{nk}. In the second iteration, we fix the r_{nk} found in the previous step and minimize J with respect to the μ_k. These two steps are repeated until convergence.

For the first stage, since J is a linear function of r_{nk} and the terms for each n are independent, we can solve for each n by assigning it to the closest cluster center. This can be written as

$$r_{nk} = \begin{cases} 1, & \text{if } k = \arg\min_{j} \|\theta_n - \mu_j\|^2 \\ 0, & \text{otherwise.} \end{cases} \tag{22}$$

In the second stage, holding the r_{nk} fixed, the objective function is a quadratic function of μ_k, and its minimum can be determined at the points where the derivative with respect to μ_k is equal to zero

$$2\sum_{n=1}^{N} r_{nk}(\theta_n - \mu_k) = 0$$

which can be easily solved as follows

$$\mu_k = \frac{\sum_n r_{nk}\theta_n}{\sum_n r_{nk}}. \tag{23}$$

The denominator in (23) is the number of data points designated to cluster k. Hence, the value of μ_k that minimizes the objective function is the mean of the θ_n currently in cluster k. This is the reason why the algorithm is known as K-means.

5.1.2 Fuzzy c-means

The K-means clustering produces a “hard” partitioning of the data set, i.e., each data point θ_n is assigned only to cluster k by setting $r_{nk} = 1$, such that $\sum_{k=1}^{K} r_{nk} = 1,\ n = 1, \ldots, N$. On the other hand, the Fuzzy c-means (FCM) algorithm allows each data point to have a membership degree associated with every cluster. These memberships span from zero to one. Namely, memberships close to zero mean little similarity between the sample and a cluster, and memberships close to one imply a high degree of similarity between the sample and that cluster (BEZDEK; EHRLICH; FULL, 1984).

For consistency's sake, we use the same symbology from Section 5.1.1. Now, let us consider the problem of a soft partitioning of a data set with N samples into K clusters, i.e., the problem of computing the optimal values for the memberships of each data point in relation to every cluster. To achieve that, the binary variables $\{r_{nk}\}$ of the K-means algorithm are replaced by a set of normalized continuous variables $\{\rho_{nk}\} \in [0,1]$, such that $\sum_{k=1}^{K} \rho_{nk} = 1,\ n = 1, \ldots, N$.

Hence, the FCM algorithm is founded on the minimization of J_m, the objective function given by

$$J_m = \sum_{n=1}^{N}\sum_{k=1}^{K} (\rho_{nk})^m \|\theta_n - \mu_k\|^2 \tag{24}$$

where θ_n is one of the N samples in the data set, μ_k is the center from one of the K clusters, and $m > 1$ is the constant that controls how fuzzy are the boundaries between cluster, such that greater values of m means a higher degree of cluster overlapping. Notice that the objective function in (24) is similar to that from K-means algorithm, presented in (21), they would be the same if we considered $m = 1$.

In our experiments, we used MATLAB implementation (THE MATHWORKS, 2022) to solve the optimization problem from the FCM algorithm. This implementation sets $m = 2$ as default and follows a two-stage iterative procedure, similar to the one applied in K-means.

Initially, the membership values $\{\rho_{nk}\}$ are defined randomly. Then, the first step is to compute the cluster centers as shown in (25).

$$\mu_k = \frac{\sum_n (\rho_{nk})^m \theta_n}{\sum_n (\rho_{nk})^m} \tag{25}$$

The second step is to calculate new values for the membership degrees following the expression in (26).

$$\rho_{nk} = \left(\sum_{i=1}^{K} \left(\frac{\|\theta_n - \mu_k\|}{\|\theta_n - \mu_i\|} \right)^{\dfrac{2}{m-1}} \right)^{-1} \tag{26}$$

With the values obtained from (25) and (26), we compute the new value for the objective function, J_m. These two steps are repeated until J_m stops improving or a maximum number of iterations is reached.

5.1.3 Spectral Clustering

Spectral clustering is a graph-based technique used to divide a data set into a given number of partitions. It has this name due to the core role that the graph spectrum plays in this approach. In his work (VON LUXBURG, 2007), von Luxburg states that spectral clustering has a simple implementation, can be solved in an efficient way, and has many fundamental advantages when compared to “traditional algorithms”. Rodriguez et al. (2019) compare a wide variety of clustering algorithms and conclude that spectral clustering has the best performance if you use the default parameters, which is a common practice among non-expert machine learning researchers.

Besides that, our main interest in that technique is not to cluster data based on graphs but to directly partition the nodes of a network. In Algorithm 3, we present the algorithm considered in our work, which is grounded on the one presented in (VON LUXBURG, 2007). Further in the text, to distinguish this algorithm from the others, we may refer to it as *K-means Spectral Clustering*.

Algorithm 3 K-means Spectral Clustering

```
Inputs: Similarity matrix $S \in \mathbb{R}^{n \times n}$, number $K$ of clusters to construct.

    Construct a similarity graph based on $k$NN[1].
    $\mathcal{W} = [w_{ij}] \in \mathbb{R}^{N \times N}$ is its weighted adjacency matrix.

    $d(i) = \sum_{j=1}^{N} w_{ij}$ is the degree of node $i$.

    $D = \mathrm{diag}([d(1) \cdots d(N)])$ is a diagonal matrix holding the degrees of every node.

    Compute the Laplacian $\mathcal{L} = D - \mathcal{W}$.

    Compute the first $K$ eigenvectors $u_1, \ldots, u_K$ of $\mathcal{L}$.

    Let $\mathcal{U} \in \mathbb{R}^{N \times K}$ be the matrix containing the vectors $u_1, \ldots, u_K$ as columns.

for $i = 1, \ldots, N$ do
    Let $y_i \in \mathbb{R}^K$ be the vector corresponding to the $i$-th row of $\mathcal{U}$.
end for

    Cluster the points $\{y_1, \ldots, y_N\}$ with the $K$-means algorithm into clusters $C_1, \ldots, C_K$.

Output: Clusters of nodes $\mathcal{V}_1, \ldots, \mathcal{V}_K$ with $\mathcal{V}_k = \{i | y_i \in C_k\}$.
```

5.2 COMMUNITY IDENTIFICATION ALGORITHM

After presenting the background on the subject, this section brings our proposed algorithm for performing community identification over complex networks. Here, the problem is approached with a soft clustering technique. The main differences from the K-means spectral clustering presented in Section 5.1.3 are that: we work directly with the graph of interpersonal relations instead of building a kNN-based similarity graph; and we use the Fuzzy c-means (FCM) algorithm to perform a soft clustering of the eigenvectors instead of using the K-means. We call this approach *FCM Spectral Clustering*.

As we have mentioned previously, for some network structures, the output of K-means spectral clustering presents errors like nodes isolated from their clusters. Aiming to mitigate this, we propose a post-processing correction stage for the algorithm, which has a relatively low computational cost. Our proposal is to perform a 2-hops neighborhood check after defuzzifying the clustering output, i.e., check if each node i in cluster C_k has at least one neighbor, j, in cluster C_k, and if j also has at least one

---

[1] k Nearest Neighbors (kNN) is a classification algorithm that assigns a data point to the more frequent class among k data points most similar to it, where k is an arbitrary parameter. The similarity graph based on kNN will connect each data point i to k other data points, in such a way that these data points have the k highest correspondent values in the i-th row of the similarity matrix S.

neighbor $l \neq i$ in cluster C_k. If that is not the case, these nodes are reallocated to the clusters with their respective second highest membership values.

We present the whole procedure in Algorithm 4. To put it in a few words: first, we apply a soft clustering technique (FCM) on the eigenvectors associated with the second and third smaller eigenvalues of the graph's Laplacian matrix; this will give each node a measure of its membership level related to each cluster; second, we defuzzify that by selecting the nodes for each cluster based on their highest membership levels; finally, we analyze whether a node (or a pair of nodes) was assigned to a cluster in which it does not have any link to the other members of the cluster; if this is the case, this node (or these nodes) is reallocated to the cluster with the second highest membership level. We call this algorithm FCM Spectral Clustering with 2 Hops Neighborhood Correction or, for short, *FCM-2Hops Spectral Clustering*.

5.3 SIMULATIONS FOR SPREAD CONTROL IN A NETWORK

We aim to maximize the spread of information in the network, i.e., reduce the time needed for a information to reach all nodes. To this purpose, we suggest combining a centrality metric with the output of our clustering algorithm. In past works from our research group (CARVALHO; MORENO, 2021; SALEM et al., 2020), we show that the eigenvector centrality is intimately related with the node's potential for spreading information.

We propose two hypothesis to be tested for selecting the initial spreaders, those who will start with the piece of information to be spread. The first one is to compute the eigenvector centrality for the entire graph, $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, and from each cluster we choose the more central node as a spreader.

The second scenario would be to construct sub-graphs containing only the nodes from each cluster, $\mathcal{G}_k = \{\mathcal{V}_k, \mathcal{E}_k\}$, where $\mathcal{V}_k = \{i | y_i \in \tilde{C}_k\}$ and $\mathcal{E}_k = \{(i,j) \in \mathcal{V}_k \times \mathcal{V}_k | (i,j) \in \mathcal{V}\}$. Then, we compute the eigenvector centrality of $\mathcal{G}_k$ and select the nodes with highest values as spreaders. This will be the policy applied on the illustrative case presented in Section 5.3.1. The data that led to choose this policy will be analyzed in Section 5.3.2.

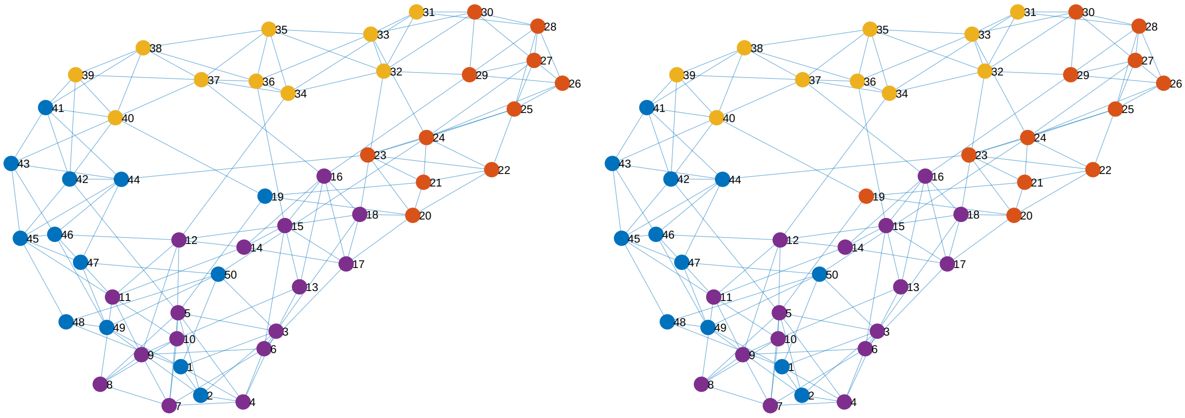
5.3.1 An Illustrative Example

In this section, we present an example application of the FCM-2Hops Spectral Clustering algorithm and one way that its output can be used to devise policies for manipulating the information spreading over the network. First, let us consider a network of 50 agents, generated according to the Watts-Strogatz model.

We perform the clustering of these agents into 4 communities, identified by the 4 colors used in Figure 49. Notice that, in Figure 49a, node 19 is assigned to the

Algorithm 4 FCM-2Hops Spectral Clustering

```
Inputs: Graph $\mathcal{G} = \{\mathcal{V},\mathcal{E}\}$, number $K$ of clusters to construct.

  Let $|\mathcal{V}| = N$ be the number of nodes in the network.
  $\mathcal{W} = [w_{ij}] \in \mathbb{R}^{N \times N}$ is its weighted adjacency matrix.

  $d(i) = \sum_{j=1}^{N} w_{ij}$ is the degree of node $i$.

  $D = \mathrm{diag}([d(1) \cdots d(N)])$ is a diagonal matrix holding the degrees of every node.

  Compute the Laplacian $\mathcal{L} = D - \mathcal{W}$.

  Compute the first $K$ eigenvectors $u_1, \ldots, u_K$ of $\mathcal{L}$.

  Let $\mathcal{U} \in \mathbb{R}^{N \times K}$ be the matrix containing the vectors $u_1, \ldots, u_K$ as columns.

for $i = 1, \ldots, N$ do
    Let $y_i \in \mathbb{R}^K$ be the vector corresponding to the $i$-th row of $\mathcal{U}$.
end for

  Use the FCM algorithm to compute the membership values $\{\rho_{ik}\}$, $i = 1, \ldots, N$, $k = 1, \ldots, K$.
  Assign the points $\{y_1, \ldots, y_N\}$ into clusters $\tilde{C}_1, \ldots, \tilde{C}_K$ by finding the cluster $\tilde{C}_k$ that has the highest membership value, $\rho_{ik}$, for each point $y_i$. Namely,
  $r_{ik} = \begin{cases} 1, & \rho_{ik} > \rho_{ip} \; \forall p \neq k \\ 0, & \text{otherwise.} \end{cases}$
Correction Stage for Node Isolation Errors:
  Let $\mathcal{N}_{ik} = \{j | y_j \in \tilde{C}_k \wedge (i,j) \in \mathcal{E}\}$ be the set of neighbors of $i$ that belongs to cluster $\tilde{C}_k$.

if $y_i \in \tilde{C}_k$ and $|\mathcal{N}_{ik}| = 0$ then
    Reassign the point $y_i$ to the cluster $\tilde{C}_\eta$ that has the second highest membership value, $\rho_{i\eta}$.
else if $y_i \in \tilde{C}_k \wedge |\mathcal{N}_{ik}| = 1 \wedge j \in \mathcal{N}_{ik} \wedge |\mathcal{N}_{jk}| = 1$ then
    Reassign points $y_i$,$y_j$ to the clusters $\tilde{C}_\eta, \tilde{C}_\nu$ that have the correspondents second highest membership values, $\rho_{i\eta}$,$\rho_{j\nu}$.
end if

Output: Clusters of nodes $\mathcal{V}_1, \ldots, \mathcal{V}_K$ with $\mathcal{V}_k = \{i | y_i \in \tilde{C}_k\}$.
```



(a) Before the correction step

(b) After the correction step

Figure 49 – Example of applying the 2-Hops Neighborhood Correction Stage to the output of the FCM Spectral Clustering.

blue cluster, even though it does not have any links to other nodes from this cluster. To amend this, we perform the correction step based on the 2 hops neighborhood explained in the previous section. The corrected clustering is depicted in Figure 49b. It is needed to highlight that the position of the nodes in these graphical representations is far less important than the links between them. For this reason, some nodes (e.g. 1, 2, and 50) may look like a bad clustering output at a first glance but analyzing their connections will help to understand why they were assigned to those clusters.

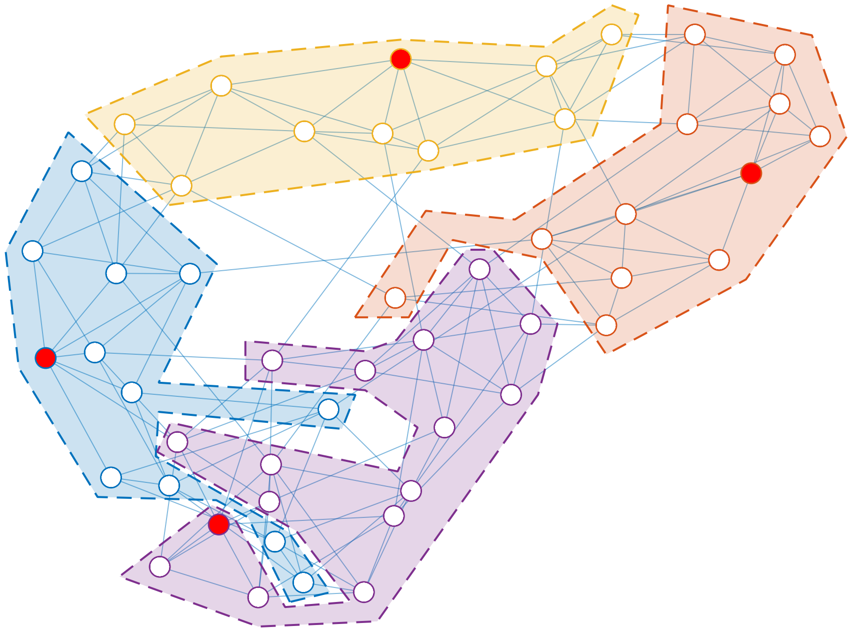
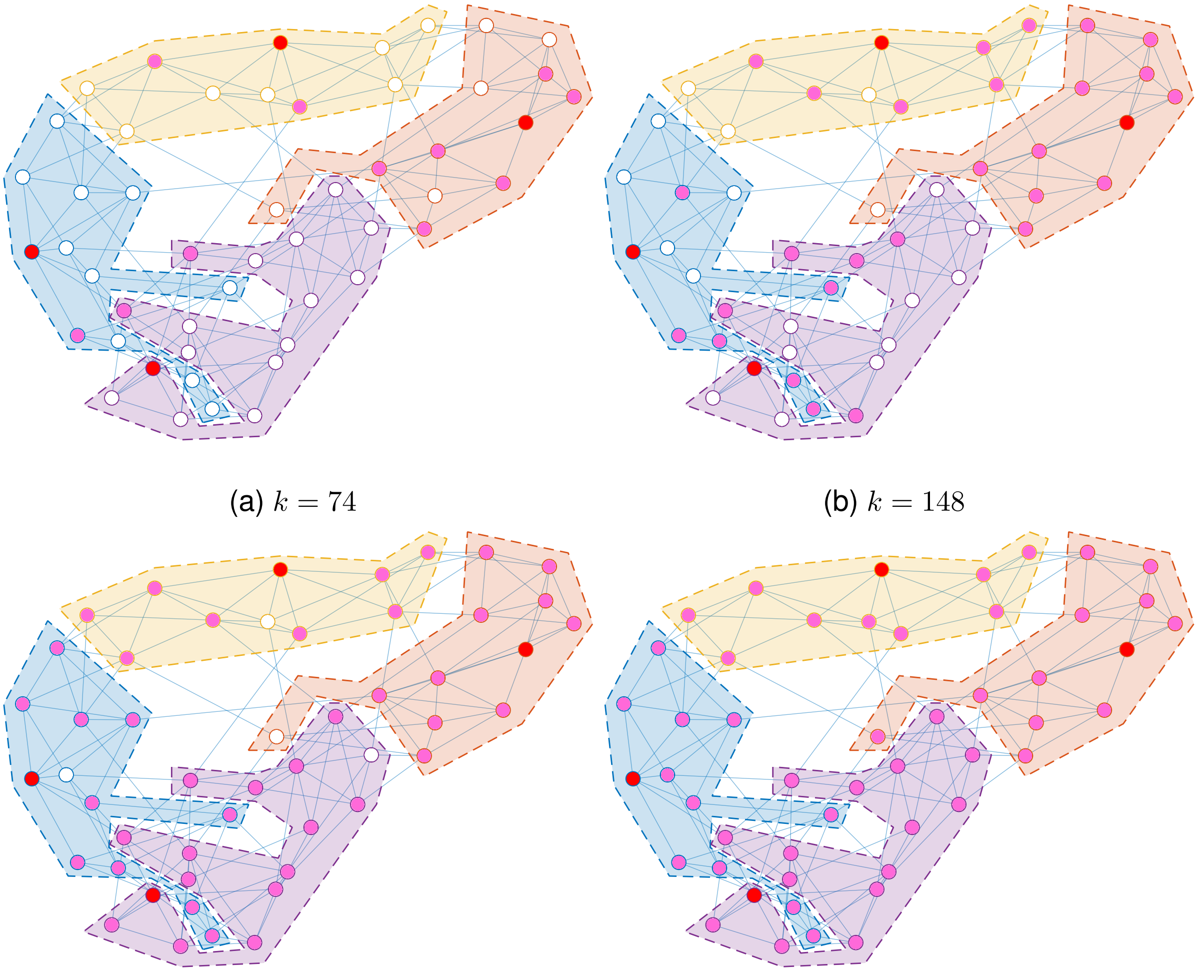


Figure 50 – Initial condition of the information spreading process

After identifying the communities, we want to apply a policy for selecting the best nodes for spreading a given piece of information. This is achieved by building 4 subgraphs, one for each cluster, and computing the eigenvector centrality for the agents in these subgraphs. Finally, the agent with the highest centrality score within its cluster will be one of the initial spreaders. To better illustrate the spreading process, in Figure 50, we have omitted the nodes' indices and painted in white the nodes which have not received the information of interest yet. The initial spreaders are identified in red and the communities are bounded by a colored hull for best visualization.


(a) $k = 74$

(b) $k = 148$

(c) $k = 221$

(d) $k = 294$

Figure 51 – States of the information spreading on the network at sample k.

Once selected the initial spreaders, the agents starts exchanging information following our model presented in Section 4.1, considering equal probabilities of interaction for every link of any given node. In this specific scenario, it took 294 samples for the information to reach all agents in the network. In Figure 51, we present 4 moments in the evolution of this spreading process. Through the QR Code in Figure 52, it is possible to watch the animation with ever iteration of this simulation. The agents represented by a white node have not received the information yet, while those represented by a pink node have already received it. The initial spreaders are still identified by a red node to differentiate them from the others as the information spreads in the network. Further considerations on the performance of the FCM-2Hops Spectral Clustering algorithm and on the suitability of the proposed policies for controlling the spread will be presented in the comparative tests of Section 5.3.2.

Figure 52 – QR Code directing for the animation illustrating the spreading evolution, also accessible through the link: https://youtu.be/7nFPHSfgzAw.

5.3.2 Comparative Analyses

In this section, we present two collections of experiments designed to evaluate the algorithm presented in Section 5.2. The goal is to measure the accuracy of the clustering algorithm, in terms of a metric formulated to account for meaningful features of this specific application. Also, we want to check if, using the knowledge acquired from this clustering technique, it is possible to construct a policy capable of making the spreading process faster.

The first set of simulations was for addressing the community identification problem. Initially, we apply the K-means spectral clustering directly over the interpersonal relations graph. In doing so, we have identified two situations that seemed to be a bad allocation by the clustering algorithm.

To be concise, we call those situations clustering errors and classify them into two types:

1. Type 1 - Isolated node: Node i is assigned to cluster k but it does not have any link connecting to another node assigned to cluster k. Namely, $i \in \mathcal{V}_k$ and $|\mathcal{N}_{ik}| = 0$, where $\mathcal{N}_{ik} = \{l \in \mathcal{V}_k | (i,l) \in \mathcal{E}\}$;
2. Type 2 - Isolated pair: Node i is assigned to cluster k but it has only one neighbor, j, in cluster k, and this relation is reciprocal. Namely, $\{i,j\} \in \mathcal{V}_k$, $(i,j) \in \mathcal{E}$, and $|\mathcal{N}_{ik}| = |\mathcal{N}_{jk}| = 1$, where $\mathcal{N}_{ik} = \{l \in \mathcal{V}_k | (i,l) \in \mathcal{E}\}$.

To exemplify these errors, we present Figures 53 and 54. Namely, in Figure 53, we depict a type 1 error, where node 31 does not have any links to the other members of its cluster. In Figure 54, we show a type 2 error, such that nodes 10 and 12 are isolated from their cluster. To better visualize this last case, we provide a magnification of that region of the graph in Figure 55.

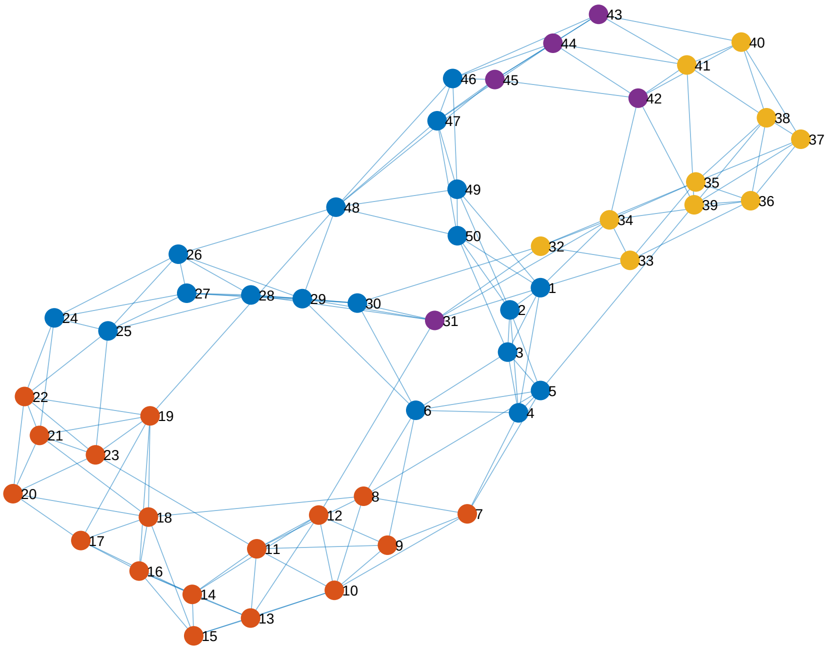


Figure 53 – Output of the K-means spectral clustering, type 1 error present at node 31

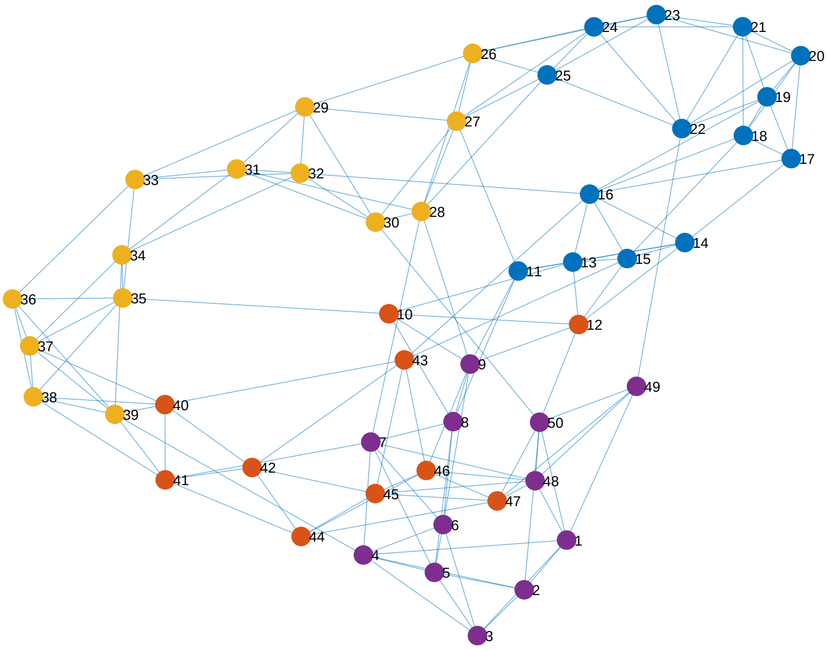


Figure 54 – Output of the K-means spectral clustering, presenting type 2 error at nodes 10 and 12

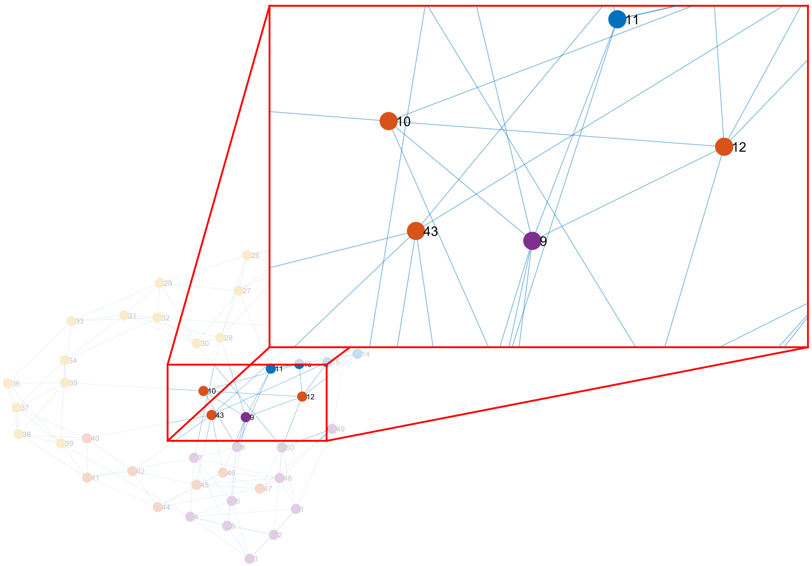


Figure 55 – Zooming into nodes 10 and 12 makes clearer that they are not connected to any other nodes from their cluster

By applying the soft clustering approach, we achieve better results in some of the analyzed network structures. This is the case shown in Figure 56, where the type 1 error present in Figure 53 was corrected. Unfortunately, the FCM spectral clustering is not able to avoid the error depicted in Figure 54, as we demonstrate in Figure 57.

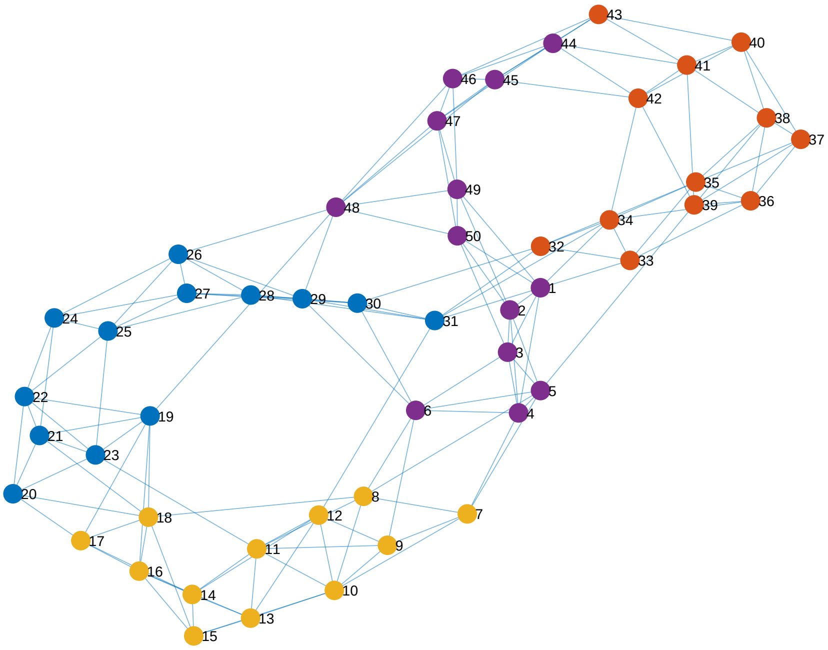


Figure 56 – Output of the FCM spectral clustering, not presenting the same error from K-means

In Figure 58, we show that the error is successfully avoided by the 2 hops correction procedure introduced in Section 5.2. However, this is an anecdotal evidence

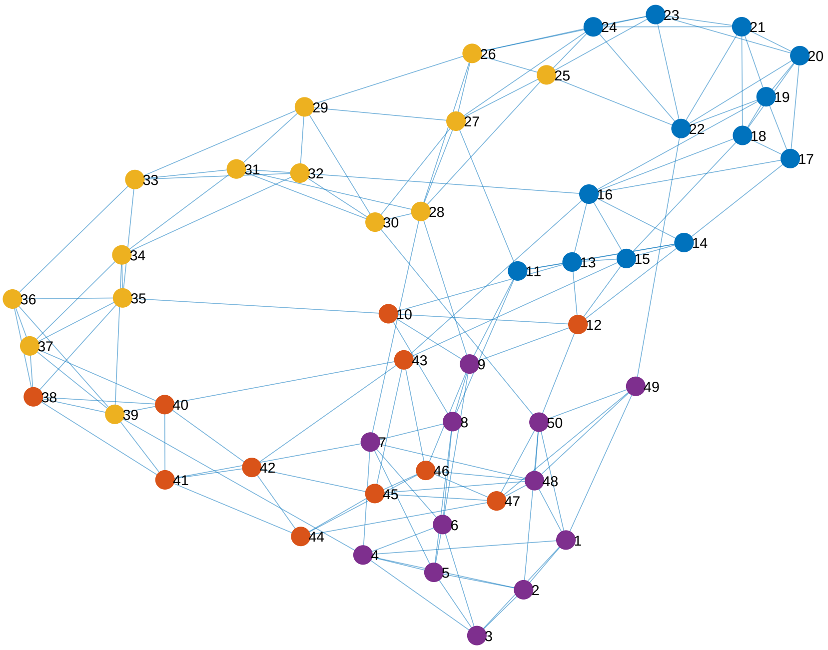


Figure 57 – Output of the FCM spectral clustering, the type 2 error persists at nodes 10 and 12

and just for purposes of illustrating the concept. Hence, a more extensive study is done and presented next.

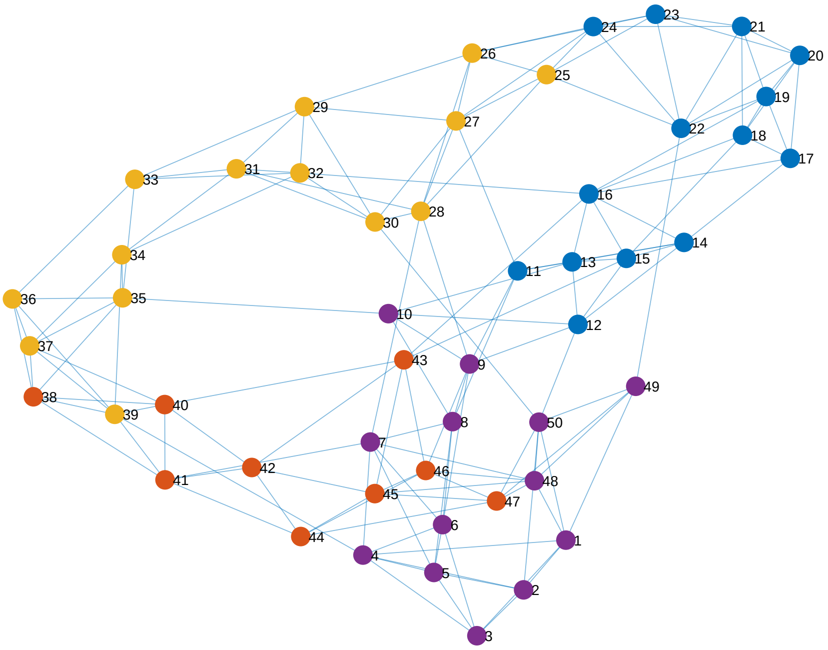


Figure 58 – Output of the FCM-2Hops spectral clustering, not presenting the same error as in the previous approaches

For comparing the 3 approaches (K-means, FCM, and FCM-2Hops), we run 1000 simulations, generating different structures in each run. These structures are the same for every algorithm and they follow the Watts-Strogatz model for small-world networks.

Also, we test the performance for 3 different sizes of networks, with 50, 100, and 500 nodes. After performing the clustering in each network, we count how many of the simulations presented errors for each algorithm. We summarize these results in Table 8.

Table 8 – Number of simulations presenting errors of any type in the clustering output for each algorithm

		Number of Nodes		
		50	100	500
Algorithm	K-Means	5	19	138
	FCM	1	21	388
	FCM-2Hops	0	0	17

From the results, we observe that the number of errors by the FCM approach increases with the network size. However the soft clustering nature of this algorithm is what allow us to perform the proposed correction step.

We notice that the number of erroneous clustering outputs drops drastically for the FCM-2Hops algorithm. For the smaller networks (50 and 100 nodes) it is able to eliminate the errors in all cases. For the 500 nodes networks, the errors are avoided in 98.3% of the simulations, while the K-means approach only gets it right in 86.2% of the cases.

The second part of this experiment has the goal of evaluating whether the proposed policies are able to make the spreading process faster. Once again, we consider networks with 50, 100, and 500 nodes and generate 1000 different structures for each size. We evaluate the two policies based on the community identification and a third one, used as a benchmark.

Considering K clusters, the three examined policies for choosing the spreaders are the following:

- **FG** - This is the benchmark, consists on simply selecting the K nodes with highest centrality values in the Full Graph;
- **CL** - The centrality values are computed for the whole graph but we select the node with the highest centrality in each Cluster;
- **SG** - The centrality values are calculated separately in each Sub-Graph, which is built by isolating the clusters. We select the more central node from each Sub-Graph.

For a more formal definition of the policies, see Section 5.3.

At the beginning of the simulations, only the spreaders have the information. For each policy application, we register how many samples are needed so the information reaches every node in the network. At first, we keep the number of cluster fixed ($K = 4$) and change the network size. The results of this trial is given in Table 9.

Table 9 – Number of samples until the information reaches all nodes (averaged over 1000 simulations), considering 4 clusters and varying the network size. Comparison between the different policies for choosing the spreaders.

Nodes	FG	CL	SG
50	537	454	415
100	1417	1265	1203
500	10860	10110	10253

For a second trial, we fix the number of nodes at 500 and test for different numbers of clusters. It is important to highlight that the number of spreaders is always equal to the number of clusters. Table 10 brings the data for comparing the 3 policies in this case.

Table 10 – Number of samples until the information reaches all nodes (averaged over 1000 simulations), fixing the network size as 500 nodes and varying the number of clusters/spreaders. Comparison between the different policies for choosing the spreaders.

Clusters	FG	CL	SG
4	10860	10110	10253
5	10692	9938	9862
6	10516	9691	9533
10	9981	9002	8663
40	8225	7191	6336

From the results of both trials, it is noticeable that the SG policy showed the best performance on average. The CL approach have also performed better than the benchmark and it was even faster than the SG approach for the scenario with 500 nodes divided into 4 clusters.

These simulations provide a preliminary evidence that our hypothesis of policy for selecting the spreader is efficient. Generically, we could say that the idea, for both CL and SG policies, is to identify the communities in a social network and pick the most central node from each community.

However, we see some limitations on the comparison conducted, which should be addressed in future works. For instance, we only evaluate for the Watts-Strogatz class of networks, which is a good and widely used representation but presents nodes with statistically similar characteristics. Analyzing other kinds of structures would strengthen our hypothesis.

Also, maybe the FG policy is not the best benchmark to be considered. The most central nodes of a network can be very close to each other. If this is the case, selecting them as the spreaders would imply in a slower spreading compared to a case when the information is being sent from different regions of the network. In any case, it is important to emphasize that this benchmark was chosen due to our previous knowledge that: if we have only one spreader, selecting the most central node would be the fastest way to reach all nodes in the network.

6 CONCLUSIONS

The main goal of this research was to propose a model that encompass the agents' non-deterministic behavior and the complex structure of the network connecting them. The core motivations for this proposal are the problems of predicting how information spreads in online social networks and the evolution of an infectious disease epidemic in a susceptible population. Improving the predictions for both cases is vital for evaluating actions that improve or counteract the spreading. For instance, it is crucial to improve the reach of prevention information during an infectious disease outbreak. By the other hand, it is desirable to counteract the spreading of “fake news” and infectious disease. These two problems can be congregated in a more generalized one, which we call spreading processes prediction.

The PhD Thesis is only a contour of all the research, study and labor put into a PhD course. We have presented some of the works elaborated during the past years, which have as it main concern the interactions between agents of a connected social system. From the wide set of simulations portrayed in this document, we can say that our models improve and widen the capacity of analysis when compared to those coming from the literature. Also, the knowledge provided by these enhanced analyses is a great tool for synthesizing network structures to control spreading processes on techno-social systems.

We have provided a model for social interactions between agents and extended it for two main case studies: information being exchanged by *online* social interactions; and diseases infecting people through *real-world* social interactions. We have built a model for the infection evolution, which is flexible to encompass different diseases and levels of detailing. This model admits probabilistic transitions between the infection stages that can advantage on (and be constantly updated by) real-world statistical data.

The model proposed for information exchange deals with finite sets of information, which is a novelty in relation to the models in the literature on opinion dynamics. The spreading in our model is dictated by the probabilities of interaction's matrix that is directly related to the network structure. However, we have identified other characteristics that impacts the information diffusion in a social network. Namely, we have modeled and incorporated the information reliability and the agents' self-esteem as factors that dynamically changes those matrix of interaction probabilities.

Since the Online Social Networks are really large and complex, any attempt to visually extract features of it would be unfruitful. Hence, we had to place our trust in statistical metrics used to describe these kind of structures. We have concluded that the eigenvector centrality is intimately related to the spreading speed and the probability of a given information prevailing over the others in our model.

For both case studies, we have provided hypothetical network structures for a variety of scenarios. Complex network models, such as the Watts-Strogatz model for Small-World networks, are better approaches to depict the topology behind our social interactions than regular or fully random networks. Therefore, all of our proposed structures rely to some extent on these models.

The most recent contribution was the clustering algorithm to be applied for identifying communities in social networks. Our proposed algorithm is able to avoid errors in the clustering that we have identified while using algorithms from the literature. We also have showed that knowing the communities splitting of a social network, and some centrality metrics about the agents, can be used to control the spreading of information in OSNs.

To summarize, we recall the specific objectives of this thesis and connect it to the sections of the document where they were developed:

- *Develop a model for the social interactions between interconnected agents* - The model proposed is presented in Section 2.4;
- *Modify an information exchange model for handling finite sets of information* - This was achieved by the model shown in Section 4.1;
- *Propose network structures that represent the scenarios of interest, founded on models of complex networks from the literature* - We have based many of our network structures on the Watts-Strogatz model for complex networks, presented in Section 2.2. And we have provided meaningful modelings for each simulation scenario, as it can be seen in Sections 3.3.2, 3.4.2, 4.1.4, 4.1.2, and 5.3.1;
- *Identify relevant features that impact the information spreading beside the network structure* - We have identified many features, but modeled and explored further the humans' self-esteem (Section 4.1.1) and the information's reliability (Section 4.1.3);
- *Build models that allow dynamic matrices of interaction probability to incorporate the identified features* - The models shown in Sections 4.1.1 and 4.1.3 have their matrices of interaction probability changing dynamically over time;
- *Construct a model for the spreading of communicable diseases based on social interactions* - The proposed model is described in Section 3.2, with close connection to the social interaction model from Section 2.4;
- *Create a stochastic model for the evolution of diseases in each individual that can use real-world data to adjust its parameters* - We propose a model for the transition between states of a disease on each individual in Section 3.2.1. In this model the probabilities of transitioning on each day can be drawn from historical data;

- *Simulate all the proposed models in meaningful scenarios to assess their capability of improving the analyses possibilities* - Simulation results and analyses are provided for all the proposed models, see Sections 3.3.3, 3.4.3, 4.1.2, 4.1.4, and 4.2.2;
- *Study characteristics from the network structure that may influence the spreading processes* - We have found that the eigenvector centrality score of a node is intimately related with the probability of its information becoming prevalent in the network, as shown in Section 4.2;
- *Adapt a clustering algorithm to identify communities in social networks* - We have built an algorithm for that purpose in Section 5.2;
- *Advantage on the knowledge acqruired from these characteristics to propose policies for controlling the spreading* - In Section 5.3, we show how the information gathered from our community identification algorithm, together with the centrality scores, can be used to improve our counteract a spreading in the network.

Furthermore, we believe that interdisciplinary collaborations can provide fundamental insights and knowledge for modeling the agent's behaviors. Also, such a contemporaneous and thriving research topic raises several ethical and even legal concerns. In order to address both information and diseases spreading, the cooperation with researchers from fields such as Marketing (business administration), Sociology, Psychology, Law, and Epidemiology (public health) should be very fruitful.

6.1 CONTRIBUTIONS

The seminal contribution of this thesis is a model for simulating social interactions between agents in a network. This model is introduced in Chapter 2 and it is used in all of the papers written during the PhD course. Departing from this common point, we can divide the upcoming contributions by the two case studies analyzed in this research: the spread of diseases in susceptible populations and the information spreading in online social networks.

On the first case study, we presented our contributions in Chapter 3, corresponding to a model for spreading of communicable diseases that considers the network governing real world social interaction between persons. Also, we built a stochastic model for the disease's evolution in each individual, based on time-dependent probabilities of passing from one stage of the disease to another. These contributions are published in the paper by Salem and Moreno (2022).

Following the case study on online social networks, we developed an information exchange model based on the asynchronous gossip algorithm, which is able to deal with a finite set of elements, i.e., not generating new pieces of information in the network through some averaging method. We presented that contribution and others in Chapter 4, and it first appeared in the paper by Salem, Tchilian, and Moreno (2019). Another contribution, introduced in the mentioned publication, is extending the models to accept dynamic matrices of interaction probabilities to capture other aspects of the information spreading process through social networks. At that point, the modeling presented considered the effect of human self-esteem on the spreading of information, while interacting with robotic nodes in OSNs groups.

Later, we continued to explore other phenomena that may impact the process of information spreading in social networks. Hence, we presented the modeling for the influence of information's reliability in the exchange through the network. This contribution was introduced in the paper by Salem, Moreno, and Lamnabhi–Lagarrigue (2020). Due to the non-deterministic nature of our models, we have explored some characteristics of the network structure to understand their behavior and get some predictions based on probabilities. In the paper by Salem et al. (2020), a contribution to the analysis of these models is presented, relating centrality metrics from the network to the prevalence probability of a given piece of information.

We presented the last contributions in Chapter 5, which refers to algorithm development for identifying communities and policies proposal to control the information spreading that relies on the knowledge about these communities acquired by applying the algorithm. Those policies also take advantage on the previous conclusions about the centrality metrics. We intend to submit these last contributions to a journal soon.

6.2 LIMITATIONS

We consider that a time-varying topology should be considered when simulating some specific scenarios. Besides that, human beings have unique and extremely complex behaviors. We accept the impossibility of modeling these behaviors fully, but we posit that a probabilistic approach can provide meaningful data for this matter.

Also, the models of opinion dynamics/information diffusion are hard to validate with real world data. We need to rely on assumptions made by works from other fields of knowledge, and most of the times is difficult for us to understand these researches or get confident that we are in the right path. This could be solved by creating an interdisciplinary research group to study these problems.

We did the research on Covid-19 due to its urgency and importance to the whole society. However, by the time that we modeled the disease, very little data and information about the disease were available. This is evident by the fact that we do not consider the possibility of a person getting reinfected by the virus. Maybe in some years, it will be possible to update the model and validate it using past data.

We have bounded our network structure models to the Watts-Strogatz model. However, there are other models in the literature (such as Power Law) that have the

Small-World characteristic and could be used to represent the scenarios studied in this thesis. Also, a more general issue is the frustration about not being able to objectively validate our models or to verify the efficiency of our proposed control policies in real-world applications.

6.3 FUTURE WORKS

In a wide overview of the results presented in this thesis, it is noticeable that the multiple specific objectives designed have made it more challenging to focus on each subject. Hence, a proposal for the following outcomes of this research could be deeper studies on the individual topics to establish mathematical properties and to apply further statistical tools for analyzing the simulation results.

Although the proposed model for epidemics was designed to allow using statistical data to adjust its probability distributions, it lacks an illustrative example of how this process should be done. Therefore, a possible unfolding of this research would be gathering more recent and consolidated data from the evolution of Covid-19 infections in the patients and demonstrating the entire pipeline, starting from this data until the probabilities in the format accepted by our model.

Another possible improvement for the epidemiology model is to use “online” data to update the predictions. This means that whenever a good measurement of the real-world status of the spreading is available it could be used to steer the model’s prediction to the right direction. This correction would be useful to avoid an unbounded growth in the variance of the model’s output as the prediction horizon reaches a long time span.

Furthermore, we envision other phenomena that influence spreading processes on Online Social Networks and could be embodied in the model. The nudge concept, popularized by Thaler and Sunstein (2009) in their book, is relevant for the information spreading problem. This theory states that human decision-making processes are highly context-dependent, and minor changes in the decision environment (called nudges) can guide individuals to a predefined choice option.

Contemporary lifestyle encompass people making increasingly more decisions on digital environments, ranging from all sorts of purchases to encountering the right life partner. This leads to a increase in relevance of the nudging concept in the digital sphere. The internet has more information than most people can handle, leading to individuals making choices in an automated way, what can make nudging easier (MIRSCH; LEHRER; JUNG, 2017). Moreover, it raises ethical questions over the use of individuals’ personal data to manipulate their decision making processes.

Also, personalization algorithms can be used to direct a more effective nudge to a specific profile of individuals. The effectiveness of digital nudging for influencing users to share contents online is subject for the research from Ni Huang et al. (2018). They presented guidelines for designing nudges that make users more willing to share a content to their friends on social networks.

Another interesting point for studies is the personality traits that could influence in the acceptance/spreading of a given information. In the Psychology field, there is a widely used model known as Big Five or OCEAN (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism) that is used to describe a person's personality as a five-dimensional variable. It could be used as a new way for setting the dynamic interactions' probability matrix present in our model.

Collecting real-world data from online social networks is also a good idea for future works, since some ideas for validating the models could arise from analyzing these data. The main challenge will be to get the data from private companies that have no interest at all in someone cracking open the shell over their algorithms.

An idea for one more interdisciplinary work is to provide data from simulations to support proposals on how to legally regulate the use of social media. Currently, there is no legal way to measure the liability of a person that spreads some malicious or false information. We believe that this liability should be judged proportionally to the person's reach in the medium used for disseminating the message. Measuring a node's capability of spreading in the network is intimately related to the work done in this thesis.

6.4 PUBLICATIONS

Following, we present the list of papers published during the PhD course. They are 3 conference papers, 1 online article for dissemination of science, and 1 journal article:

- SALEM, Feres Azevedo; TCHILIAN, Renan da Silva; MORENO, Ubirajara Franco. Evolution of Discrete Opinions on Human-Swarm Interaction in Influence Networks. In: ANAIS DO 14º SIMPÓSIO BRASILEIRO DE AUTOMAÇÃO INTELIGENTE, 2019, Ouro Preto. ANAIS Eletrônicos... Ouro Preto: Galoá, 2019
- SALEM, Feres Azevedo et al. Opinion Dynamics over a Finite Set in Cooperative Multi-robot Systems: An Asynchronous Gossip-Based Consensus Approach. In: ANAIS DO XXIII CONGRESSO BRASILEIRO DE AUTOMÁTICA, 2020, Porto Alegre. ANAIS Eletrônicos... Porto Alegre: Galoá, 2020. In Press
- SALEM, Feres Azevedo; MORENO, Ubirajara Franco; LAMNABHI–LAGARRIGUE, Françoise. A control approach to address ethical issues on social (robotic) networks. In: VIRTUAL IFAC WORLD CONGRESS, 1., 2020, Berlin. IFAC-PAPERSONLINE. Berlin: International Federation of Automatic Control (IFAC), 2020. In Press
- SALEM, Feres Azevedo; MORENO, Ubirajara Franco. Simulação estocástica: Simulações para análise do impacto da retomada de atividades presenciais

de ensino na UFSC. Florianópolis, 2020. Available at: <`https://covid19sc.github.io/simulacoes_probabilisticas.html`>. Accessed on: 26 Jun. 2020

- SALEM, Feres Azevedo; MORENO, Ubirajara Franco. A Multi-Agent-Based Simulation Model for the Spreading of Diseases Through Social Interactions During Pandemics. **Journal of Control, Automation and Electrical Systems**, Springer, p. 1–16, 2022

REFERENCES

ALMEIDA, Michelle Lau de; MORGADO, Igor; ARENAS, Zochil González. Modelo SIR Estocástico na Transmissão da COVID-19 no Rio de Janeiro - Abordagem preliminar. **Anais da Sociedade Brasileira de Automática**, v. 2, n. 1, 2020. `https://doi.org/10.48011/asba.v2i1.983`.

BALTSOU, Georgia; CHRISTOPOULOS, Konstantinos; TSICHLAS, Konstantinos. Local Community Detection: A Survey. **IEEE Access**, IEEE, v. 10, p. 110701–110726, 2022.

BARABÁSI, Albert-László et al. **Network Science**. [S.l.]: Cambridge University Press, 2016.

BATISTA, Fernando C; CUNHA JR, Americo. Inferência dos Estágios Iniciais da COVID-19 em Portugal. **Anais da Sociedade Brasileira de Automática**, v. 2, n. 1, 2020. `https://doi.org/10.48011/asba.v2i1.978`.

BATISTELA, Cristiane M; CABRERA, Manuel AM, et al. COVID-19: Estudo da imunização usando modelo SIR. **Anais da Sociedade Brasileira de Automática**, v. 2, n. 1, 2020. `https://doi.org/10.48011/asba.v2i1.979`.

BERNAYS, Edward L. **Propaganda**. [S.l.]: Ig publishing, 2005.

BEZDEK, James C; EHRLICH, Robert; FULL, William. FCM: The fuzzy c-means clustering algorithm. **Computers & geosciences**, Elsevier, v. 10, n. 2-3, p. 191–203, 1984.

BISHOP, Christopher M; NASRABADI, Nasser M. **Pattern recognition and machine learning**. [S.l.]: Springer, 2006. v. 4.

BONACICH, Phillip. Technique for analyzing overlapping memberships. **Sociological methodology**, JSTOR, v. 4, p. 176–185, 1972.

BONDY, J. Adrian; MURTY, U. S. R. **Graph theory**. 3. ed. London: Springer-Verlag, 2008. v. 244, p. 663. (Graduate texts in mathematics). ISBN 978-1-84628-969-9.

BOVET, Alexandre; MAKSE, Hernán A. Influence of fake news in Twitter during the 2016 US presidential election. **Nature communications**, Nature Publishing Group, v. 10, n. 1, p. 1–14, 2019.

BOYD, Danah M.; ELLISON, Nicole B. Social Network Sites: Definition, History, and Scholarship. **Journal of Computer-Mediated Communication**, v. 13, n. 1, p. 210–230, 2007.

BOYD, Stephen; GHOSH, Arpita; PRABHAKAR, Balaji; SHAH, Devavrat. Randomized gossip algorithms. **IEEE/ACM Transactions on Networking (TON)**, IEEE Press, v. 14, SI, p. 2508–2530, 2006.

BRADSHAW, Samantha; HOWARD, Philip. Troops, trolls and troublemakers: A global inventory of organized social media manipulation. Oxford Internet Institute, 2017.

BRAUER, Fred. Compartmental models in epidemiology. In: MATHEMATICAL epidemiology. [S.l.]: Springer, 2008. `https://doi.org/10.1007/978-3-540-78911-6_2`. P. 19–79.

BULLO, Francesco. **Lectures on Network Systems**. 1.3. ed. [S.l.]: Kindle Direct Publishing, 2019. with contributions by J. Cortés, F. Dörfler, and S. Martínez. ISBN 978-1986425643.

CARVALHO, Sidney Roberto Dias de. **Network topology control for connectivity maintenance and information spreading manipulation in multi-robot systems**. 2020. PhD thesis – Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia de Automação e Sistemas, Florianópolis.

CARVALHO, Sidney Roberto Dias de; CORREIA, Filipe Lopes de Barros; MORENO, Ubirajara Franco. Topology control for connectivity maintenance in cooperative mobile robot networks. **IFAC-PapersOnLine**, Elsevier, v. 48, n. 19, p. 280–285, 2015.

CARVALHO, Sidney Roberto Dias de; MORENO, Ubirajara Franco. Optimal design for manipulation of random consensus over discrete information in networked systems. **Journal of the Franklin Institute**, Elsevier, v. 358, n. 9, p. 5039–5072, 2021.

COLLINGSWORTH, Ben; MENEZES, Ronaldo. SOCIAL: A self-organized entropy-based algorithm for identifying communities in networks. In: IEEE. 2012 IEEE Sixth International Conference on Self-Adaptive and Self-Organizing Systems. [S.l.: s.n.], 2012. P. 217–222.

COOLEY, Charles Horton. **Human nature and the social order**. [S.l.]: Routledge, 2017.

CORREIA, Filipe Lopes de Barros; MORENO, Ubirajara Franco. Decentralized Formation Tracking for Groups of Mobile Robots with Consensus and MPC. **IFAC-PapersOnLine**, Elsevier, v. 48, n. 19, p. 274–279, 2015.

COSTA JR, Jorge A; MARTINEZ, Amanda C; GEROMEL, José C. On the Continuous-time and Discrete-Time Versions of an Alternative Epidemic Model of the SIR Class. **Journal of Control, Automation and Electrical Systems**, Springer, v. 33, n. 1, p. 38–48, 2022. `https://doi.org/10.1007/s40313-021-00757-2`.

CUEVAS, Erik. An agent-based model to evaluate the COVID-19 transmission risks in facilities. **Computers in Biology and Medicine**, Elsevier, p. 103827, 2020. `https://doi.org/10.1016/j.compbiomed.2020.103827`.

DABBENE, Fabrizio; RAVAZZI, Chiara; PROSKURNIKOV, Anton V.; ANDERSON, Brian D. O.; YE, Mengbin. Dynamical Networks of Social Influence: Modern Trend and Perspectives. In: VIRTUAL IFAC 2020 World Congress. Berlim, Alemanha: [s.n.], 2020. Available at: <`https://www.ifac2020.org/program/tutorials/dynamical-networks-of-social-influence-modern-trends-and-perspectives/`>. Accessed on: 17 Jul. 2020.

DALEY, D. J.; GANI, J. **Epidemic Modelling: An Introduction**. New York, NY, USA: Cambridge University Press, 1999. (Cambridge Studies in Mathematical Biology).

DEGROOT, Morris H. Reaching a consensus. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 69, n. 345, p. 118–121, 1974.

DIAS, Samaherni; QUEIROZ, Kurios; ARAUJO, Aldayr. Controlling epidemic diseases based only on social distancing level. **Journal of Control, Automation and Electrical Systems**, Springer, v. 33, n. 1, p. 8–22, 2022. `https://doi.org/10.1007/s40313-021-00745-6`.

DIAS, Samaherni; QUEIROZ, Kurios; ARAUJO, Aldayr. Introduction to Group-Structured-Epidemic Model. **Journal of Control, Automation and Electrical Systems**, Springer, v. 33, n. 1, p. 23–37, 2022. `https://doi.org/10.1007/s40313-021-00841-7`.

DOROGOVTSEV, Sergey N.; GOLTSEV, Alexander V.; MENDES, José F. F. Critical phenomena in complex networks. **Reviews of Modern Physics**, APS, v. 80, n. 4, p. 1275, 2008.

DREWS, Martin; KUMAR, Pavan; SINGH, Ram Kumar; DE LA SEN, Manuel; SINGH, Sati Shankar; PANDEY, Ajai Kumar; KUMAR, Manoj; RANI, Meenu; SRIVASTAVA, Prashant Kumar. Model-based ensembles: Lessons learned from retrospective analysis of COVID-19 infection forecasts across 10 countries. **Science of the Total Environment**, Elsevier, v. 806, p. 150639, 2022. `https://doi.org/10.1016/j.scitotenv.2021.150639`.

EKINCI, Aykut. Modelling and forecasting of growth rate of new COVID-19 cases in top nine affected countries: Considering conditional variance and asymmetric effect. **Chaos, Solitons & Fractals**, Elsevier, v. 151, p. 111227, 2021. `https://doi.org/10.1016/j.chaos.2021.111227`.

EODS. **Compendium of international standards for elections**. [S.l.]: Publication Office of the European Union, 2016. Fourth Edition.

ESTRADA, Ernesto. COVID-19 and SARS-CoV-2. Modeling the present, looking at the future. **Physics Reports**, Elsevier, 2020. `https://doi.org/10.1016/j.physrep.2020.07.005`.

FAY, Sarah; HOSOI, Anette; JONES, Dalton. Pandemics: Modeling and Control. In: ANNASWAMY, Anuradha M.; JOHANSSON, Karl H.; PAPPAS, George J. (Eds.). **Control for Societal-Scale Challenges: Road Map 2030**. [S.l.]: IEEE Control Systems Society, 2023.

FORTUNATO, Santo. Community detection in graphs. **Physics reports**, Elsevier, v. 486, n. 3-5, p. 75–174, 2010.

FREEMAN, L.C. **The Development of Social Network Analysis: A Study in the Sociology of Science**. [S.l.]: Empirical Press, 2004. ISBN 9781594577147.

FRIEDKIN, Noah E. A formal theory of reflected appraisals in the evolution of power. **Administrative Science Quarterly**, Sage Publications Sage CA: Los Angeles, CA, v. 56, n. 4, p. 501–529, 2011.

GLEICH, David F. PageRank beyond the Web. **SIAM Review**, SIAM, v. 57, n. 3, p. 321–363, 2015.

GODSIL, Chris; ROYLE, Gordon F. **Algebraic Graph Theory**. 1. ed. New York: Springer-Verlag, 2001. v. 207, p. 443. (Graduate texts in mathematics). ISBN 978-0-387-95220-8.

GOLDBARG, Marco; GOLDBARG, Elizabeth. **Grafos: Conceitos, algoritmos e aplicações**. 1. ed. Rio de Janeiro: Elsevier, 2012. P. 640. ISBN 978-85-352-5716-8.

GOMES, Daiana Caroline dos Santos; SERRA, Ginalber Luiz de Oliveira. Computational Approach For Real-Time Interval Type-2 Fuzzy Kalman Filtering and Forecasting via Unobservable Spectral Components of Experimental Data. **Journal of Control, Automation and Electrical Systems**, Springer, v. 32, n. 2, p. 326–355, 2021. `https://doi.org/10.1007/s40313-020-00675-9`.

GRANOVETTER, Mark. The strength of weak ties: A network theory revisited. **Sociological theory**, JSTOR, p. 201–233, 1983.

GROSS, Thilo; D'LIMA, Carlos J Dommar; BLASIUS, Bernd. Epidemic dynamics on an adaptive network. **Physical review letters**, APS, v. 96, n. 20, p. 208701, 2006.

GUMEL, Abba B.; IBOI, Enahoro A.; NGONGHALA, Calistus N.; ELBASHA, Elamin H. A primer on using mathematics to understand COVID-19 dynamics: Modeling, analysis and simulations. **Infectious Disease Modelling**, Elsevier, v. 6, p. 148–168, Jan. 2021. `https://doi.org/10.1016/J.IDM.2020.11.005`. ISSN 2468-0427.

HAN, Byung-Chul. **In the swarm: Digital prospects**. [S.l.]: MIT Press, 2017. v. 3.

HE, Sha; TANG, Sanyi; RONG, Libin. A discrete stochastic model of the COVID-19 outbreak: Forecast and control. **Math. Biosci. Eng**, v. 17, p. 2792–2804, 2020. `https://doi.org/10.3934/mbe.2020153`.

HE, Xi et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. **Nature medicine**, Nature Publishing Group, v. 26, n. 5, p. 672–675, 2020. `https://doi.org/10.1038/s41591-020-0869-5`.

HETHCOTE, Herbert W. The Basic Epidemiology Models: Models, Expressions for R_0, Parameter Estimation, and Applications. In: MA, Stefan; XIA, Yingcun (Eds.). **Mathematical Understanding of Infectious Disease Dynamics**. Singapore: World Scientific Publishing Co. Pte. Ltd., 2009. v. 16. (Lecture Notes Series). P. 1–61.

HOWARD, P et al. How political campaigns weaponize social media bots. **IEEE Spectrum**, 2018.

HUANG, Ni; CHEN, Peiyu; HONG, Yili; WU, Shinyi. Digital nudging for online social sharing: Evidence from a randomized field experiment. In: PROCEEDINGS of the 51st Hawaii International Conference on System Sciences. [S.l.: s.n.], 2018.

HUANG, Rui; LIU, Miao; DING, Yongmei. Spatial-temporal distribution of COVID-19 in China and its prediction: A data-driven modeling analysis. **The Journal of Infection in Developing Countries**, v. 14, n. 03, p. 246–253, 2020. `https://doi.org/10.3855/jidc.12585`.

HUNTER, Elizabeth; MAC NAMEE, Brian; KELLEHER, John, et al. A Hybrid Agent-Based and Equation Based Model for the Spread of Infectious Diseases. **Journal of Artificial Societies and Social Simulation**, Journal of Artificial Societies and Social Simulation, v. 23, n. 4, p. 1–14, 2020. `https://doi.org/10.18564/jasss.4421`.

IFAC 2023-YOKOHAMA. **Open invited track list | IFAC 2023 - 22nd World Congress of the International Federation of Automatic Control**. [S.l.: s.n.], 2020. Available at: <`https://www.ifac2023.org/submission/open-invited-track-list/#Control%20for%20Socio-Technical%20Network%20Systems`>. Accessed on: 8 Oct. 2022.

JIA, Peng; MIRTABATABAEI, Anahita; FRIEDKIN, Noah E.; BULLO, Francesco. Opinion dynamics and the evolution of social power in influence networks. **SIAM review**, SIAM, v. 57, n. 3, p. 367–397, 2015.

KATZ, Leo. A new status index derived from sociometric analysis. **Psychometrika**, Springer, v. 18, n. 1, p. 39–43, 1953.

KERMACK, William Ogilvy; MCKENDRICK, Anderson G. A contribution to the mathematical theory of epidemics. **Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character**, The Royal Society London, v. 115, n. 772, p. 700–721, 1927. `https://doi.org/10.1098/rspa.1927.0118`.

KUCHARSKI, Adam. Study epidemiology of fake news. **Nature**, Nature Publishing Group UK London, v. 540, n. 7634, p. 525–525, 2016.

LAMNABHI–LAGARRIGUE, Françoise et al. Systems & Control for the future of humanity, research agenda: Current and future roles, impact and grand challenges. **Annual Reviews in Control**, Elsevier, v. 43, p. 1–64, 2017.

LEWANDOWSKY, Stephan; ECKER, Ullrich KH; COOK, John. Beyond misinformation: Understanding and coping with the "post-truth" era. **Journal of applied research in memory and cognition**, Elsevier, v. 6, n. 4, p. 353–369, 2017.

LI, Zhongkui; DUAN, Zhisheng. **Cooperative Control of Multi-Agent Systems**: A Consensus Region Approach. 1. ed. Boca Raton: CRC Press, 2014. P. 250. (Automation and Control Engineering). ISBN 978-1-4665-6997-3.

LIPPMANN, Walter. **Public opinion**. [S.l.]: Routledge, 2017.

LIU, Tao et al. Time-Varying Transmission Dynamics of Novel Coronavirus Pneumonia in China. **bioRxiv**, Cold Spring Harbor Laboratory, 2020.

LIU, Yang. Estimating the case fatality rate for COVID-19: a Markov model application. **Available at SSRN 3558612**, 2020.

LUCCHESI, Cláudio Leonardo. **Introdução à teoria dos grafos**. Rio de Janeiro: IMPA, 1979. P. 148.

MAHASE, Elisabeth. COVID-19: Where are we on vaccines and variants? **British Medical Journal**, BMJ Publishing Group Ltd, v. 372, 2021. `https://doi.org/10.1136/bmj.n597`.

MANDAL, Manotosh; JANA, Soovoojeet; NANDI, Swapan Kumar; KHATUA, Anupam; ADAK, Sayani; KAR, TK. A model based study on the dynamics of COVID-19:

Prediction and control. **Chaos, Solitons & Fractals**, Elsevier, p. 109889, 2020. `https://doi.org/10.1016/j.chaos.2020.109889`.

MANZO, Gianluca. Complex Social Networks are Missing in the Dominant COVID-19 Epidemic Models. **Sociologica**, v. 14, n. 1, p. 31–49, 2020.

MARFAK, Abdelghafour; ACHAK, Doha; AZIZI, Asmaa; NEJJARI, Chakib; ABOUDI, Khalid; SAAD, Elmadani; HILALI, Abderraouf; YOULYOUZ-MARFAK, Ibtissam. The hidden Markov chain modelling of the COVID-19 spreading using Moroccan dataset. **Data in brief**, Elsevier, v. 32, p. 106067, 2020.

MCLUHAN, Marshall. **Understanding media : the extensions of man**. New York: The MIT Press, 1994. Reprint Edition.

MEDEIROS, Marcelo C; STREET, Alexandre; VALLADÃO, Davi; VASCONCELOS, Gabriel; ZILBERMAN, Eduardo. Short-term Covid-19 forecast for latecomers. **International journal of forecasting**, Elsevier, 2021. `https://doi.org/10.1016/j.ijforecast.2021.09.013`.

MEDREK, M; PASTUSZAK, Z. Numerical simulation of the novel coronavirus spreading. **Expert Systems with Applications**, Elsevier, v. 166, p. 114109, 2020. `https://doi.org/10.1016/j.eswa.2020.114109`.

MESBAHI, Mehran; EGERSTEDT, Magnus. **Graph Theoretic Methods in Multiagent Networks**. 1st. Princeton, New Jersey, USA: Princeton University Press, 2010. (Applied Mathematics).

MIRSCH, Tobias; LEHRER, Christiane; JUNG, Reinhard. Digital nudging: Altering user behavior in digital environments. **Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017)**, p. 634–648, 2017.

MO, Baichuan; FENG, Kairui; SHEN, Yu; TAM, Clarence; LI, Daqing; YIN, Yafeng; ZHAO, Jinhua. Modeling epidemic spreading through public transit using time-varying encounter network. **Transportation Research Part C: Emerging Technologies**, Pergamon, v. 122, p. 102893, Jan. 2021. `https://doi.org/10.1016/j.trc.2020.102893`. ISSN 0968090X.

NEUDERT, Lisa Maria; MARCHAL, Nahema. **Polarisation and the use of technology in political campaigns and communication**. [S.l.]: European

Parliamentary Research Service, 2019. STOA - Panel for the Future of Science and Technology.

NIELSEN, Bjarke Frost; SIMONSEN, Lone; SNEPPEN, Kim. COVID-19 Superspreading Suggests Mitigation by Social Network Modulation. **Physical Review Letters**, v. 126, n. 11, 2021. `https://doi.org/10.1103/PhysRevLett.126.118301`.

ORO, Sheila R; HELLMANN, Liliane; MAFIOLETI, Tereza R; DI DOMÊNICO, Camila NB; CAMPOS, Guilherme L. Modelagem dinâmica para previsão dos casos novos de COVID-19 no Estado do Paraná. **Anais da Sociedade Brasileira de Automática**, v. 2, n. 1, 2020. `https://doi.org/10.48011/asba.v2i1.990`.

PAIVA, Henrique Mohallem; AFONSO, Rubens Junqueira Magalhães; LIMA ALVARENGA, Fabiana Mara Scarpelli de; ANDRADE VELASQUEZ, Ester de, et al. A computational tool for trend analysis and forecast of the COVID-19 pandemic. **Applied Soft Computing**, Elsevier, v. 105, p. 107289, 2021. `https://doi.org/10.1016/j.asoc.2021.107289`.

PARINO, Francesco; ZINO, Lorenzo; PORFIRI, Maurizio; RIZZO, Alessandro. Modelling and predicting the effect of social distancing and travel restrictions on COVID-19 spreading. **Journal of the Royal Society Interface**, Royal Society Publishing, v. 18, 175 Feb. 2021. `https://doi.org/10.1098/RSIF.2020.0875`. ISSN 17425662.

PARISER, Eli. **The filter bubble: What the Internet is hiding from you**. [S.l.]: penguin UK, 2011.

PARSEGOV, Sergey E.; PROSKURNIKOV, Anton V.; TEMPO, Roberto; FRIEDKIN, Noah E. Novel multidimensional models of opinion dynamics in social networks. **IEEE Transactions on Automatic Control**, IEEE, v. 62, n. 5, p. 2270–2285, 2017.

PASTOR-SATORRAS, Romualdo; CASTELLANO, Claudio; VAN MIEGHEM, Piet; VESPIGNANI, Alessandro. Epidemic processes in complex networks. **Reviews of modern physics**, APS, v. 87, n. 3, p. 925, 2015. `https://doi.org/10.1103/RevModPhys.87.925`.

PATARO, Igor ML; MORATO, Marcelo M; AMERICANO DA COSTA, Marcus V; NORMEY-RICO, Julio E. Optimal control approach for the COVID-19 pandemic in Bahia and Santa Catarina, Brazil. **Journal of Control, Automation and Electrical Systems**, Springer, v. 33, n. 1, p. 49–62, 2022. `https://doi.org/10.1007/s40313-021-00760-7`.

PAZOS, Fernando A; FELICIONI, Flavia. A control approach to the COVID-19 disease using a SEIHRD dynamical model. **Anais da Sociedade Brasileira de Automática**, v. 2, n. 1, 2020. `https://doi.org/10.48011/asba.v2i1.1002`.

PERRUQUETTI, Wilfrid. Social Impacts of Automation | Control Challenges for Social Systems Panel Session. In: IFAC 2017 World Congress. Toulouse, França: [s.n.], 2017. Available at: <`https://youtu.be/CXy8GCfRZSg?t=530`>. Accessed on: 8 Mar. 2020.

PROSKURNIKOV, Anton V; RAVAZZI, Chiara; DABBENE, Fabrizio. Dynamics and structure of social networks from a systems and control viewpoint: A survey of Roberto Tempo's contributions. **Online Social Networks and Media**, Elsevier, v. 7, p. 45–59, 2018.

PROSKURNIKOV, Anton V; TEMPO, Roberto. A tutorial on modeling and analysis of dynamic social networks. Part I. **Annual Reviews in Control**, Elsevier, v. 43, p. 65–79, 2017.

PROSKURNIKOV, Anton V; TEMPO, Roberto. A tutorial on modeling and analysis of dynamic social networks. Part II. **Annual Reviews in Control**, Elsevier, v. 45, p. 166–190, 2018.

RAMOS, A. M.; FERRÁNDEZ, M. R.; VELA-PÉREZ, M.; KUBIK, A. B.; IVORRA, B. A simple but complex enough θ-SIR type model to be used with COVID-19 real data. Application to the case of Italy. **Physica D: Nonlinear Phenomena**, North-Holland, v. 421, p. 132839, July 2021. `https://doi.org/10.1016/J.PHYSD.2020.132839`. ISSN 0167-2789.

REN, Wei; BEARD, Randal W. **Distributed Consensus in Multi-vehicle Cooperative Control**: Theory and Applications. 1. ed. London: Springer-Verlag, 2008. P. 319. (Communications and Control Engineering). ISBN 978-1-84800-014-8.

REN, Wei; BEARD, Randal W.; ATKINS, Ella M. Information consensus in multivehicle cooperative control. **IEEE Control Systems**, IEEE, v. 27, n. 2, p. 71–82, 2007.

RIBEIRO, Matheus Henrique Dal Molin; SILVA, Ramon Gomes da; MARIANI, Viviana Cocco; SANTOS COELHO, Leandro dos. Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. **Chaos, Solitons & Fractals**, Elsevier, v. 135, p. 109853, 2020. `https://doi.org/10.1016/j.chaos.2020.109853`.

RICCI, Francesco; ROKACH, Lior; SHAPIRA, Bracha. Recommender systems: introduction and challenges. **Recommender systems handbook**, Springer, p. 1–34, 2015.

ROBERTS, Mick; ANDREASEN, Viggo; LLOYD, Alun; PELLIS, Lorenzo. Nine challenges for deterministic epidemic models. **Epidemics**, Elsevier, v. 10, p. 49–53, 2015. `https://doi.org/10.1016/j.epidem.2014.09.006`.

RODRIGUEZ, Mayra Z; COMIN, Cesar H; CASANOVA, Dalcimar; BRUNO, Odemir M; AMANCIO, Diego R; COSTA, Luciano da F; RODRIGUES, Francisco A. Clustering algorithms: A comparative approach. **PloS one**, Public Library of Science San Francisco, CA USA, v. 14, n. 1, e0210236, 2019.

RUBENSTEIN, Michael; AHLER, Christian; NAGPAL, Radhika. Kilobot: A Low Cost Scalable Robot System for Collective Behaviors. In: IEEE. 2012 IEEE International Conference on Robotics and Automation. Saint Paul, MN, USA: [s.n.], 2012. P. 3293–3298.

SALEM, Feres Azevedo; MORENO, Ubirajara Franco. A Multi-Agent-Based Simulation Model for the Spreading of Diseases Through Social Interactions During Pandemics. **Journal of Control, Automation and Electrical Systems**, Springer, p. 1–16, 2022.

SALEM, Feres Azevedo; MORENO, Ubirajara Franco. Simulação estocástica: Simulações para análise do impacto da retomada de atividades presenciais de ensino na UFSC. Florianópolis, 2020. Available at: <`https://covid19sc.github.io/simulacoes_probabilisticas.html`>. Accessed on: 26 Jun. 2020.

SALEM, Feres Azevedo; MORENO, Ubirajara Franco; CASTELAN, Eugênio de Bona. Information Distributed Kalman Filter Applied to Rendezvous Problems in Cooperative Robotic Teams. **IFAC-PapersOnLine**, Elsevier, v. 51, n. 25, p. 190–195, 2018.

SALEM, Feres Azevedo; MORENO, Ubirajara Franco; LAMNABHI–LAGARRIGUE, Françoise. A control approach to address ethical issues on social (robotic) networks. In: VIRTUAL IFAC WORLD CONGRESS, 1., 2020, Berlin. IFAC-PAPERSONLINE. Berlin: International Federation of Automatic Control (IFAC), 2020. In Press.

SALEM, Feres Azevedo; TCHILIAN, Renan da Silva; CARVALHO, Sidney Roberto Dias de; MORENO, Ubirajara Franco. Opinion Dynamics over a Finite Set in Cooperative Multi-robot Systems: An Asynchronous Gossip-Based Consensus Approach. In: ANAIS DO XXIII CONGRESSO BRASILEIRO DE AUTOMÁTICA, 2020, Porto Alegre. ANAIS Eletrônicos... Porto Alegre: Galoá, 2020. In Press.

SALEM, Feres Azevedo; TCHILIAN, Renan da Silva; MORENO, Ubirajara Franco. Evolution of Discrete Opinions on Human-Swarm Interaction in Influence Networks. In: ANAIS DO 14º SIMPÓSIO BRASILEIRO DE AUTOMAÇÃO INTELIGENTE, 2019, Ouro Preto. ANAIS Eletrônicos... Ouro Preto: Galoá, 2019.

SANTOS, Débora Marques Lopes; RODRIGUES, Victor Hugo Pereira; OLIVEIRA, Tiago Roux. Epidemiological Control of COVID-19 Through the Theory of Variable Structure and Sliding Mode Systems. **Journal of Control, Automation and Electrical Systems**, Springer, v. 33, n. 1, p. 63–77, 2022. `https://doi.org/10.1007/s40313-021-00826-6`.

SCABINI, Leonardo FS; RIBAS, Lucas C; NEIVA, Mariane B; JUNIOR, Altamir GB; FARFAN, Alex JF; BRUNO, Odemir M. Social interaction layers in complex networks for the dynamical epidemic modeling of COVID-19 in Brazil. **Physica A: Statistical Mechanics and its Applications**, Elsevier, v. 564, p. 125498, 2021.

SCALA, A; FLORI, A; SPELTA, A; BRUGNOLI, E; CINELLI, M; QUATTROCIOCCHI, W; PAMMOLLI, F. Time, space and social interactions: Exit mechanisms for the COVID-19 epidemics. **Nature Scientific Reports**, Nature Publishing Group, v. 10, 2020. `https://doi.org/10.1038/s41598-020-70631-9`.

SHAO, Chengcheng; CIAMPAGLIA, Giovanni Luca; VAROL, Onur; YANG, Kai-Cheng; FLAMMINI, Alessandro; MENCZER, Filippo. The spread of low-credibility content by social bots. **Nature communications**, Nature Publishing Group, v. 9, n. 1, p. 4787, 2018.

SHI, Pengpeng; CAO, Shengli; FENG, Peihua. SEIR Transmission Dynamics Model of 2019 nCoV Coronavirus with Considering the Weak Infectious Ability and Changes in Latency Duration. **medRxiv**, Cold Spring Harbor Laboratory Press, 2020.

SILVA, Ramon Gomes da; RIBEIRO, Matheus Henrique Dal Molin; MARIANI, Viviana Cocco; SANTOS COELHO, Leandro dos. Forecasting Brazilian and American COVID-19 cases based on artificial intelligence coupled with climatic exogenous variables. **Chaos, Solitons & Fractals**, Elsevier, v. 139, p. 110027, 2020. `https://doi.org/10.1016/j.chaos.2020.110027`.

STOKES, Erin K; ZAMBRANO, Laura D; ANDERSON, Kayla N; MARDER, Ellyn P; RAZ, Kala M; FELIX, Suad El Burai; TIE, Yunfeng; FULLERTON, Kathleen E. Coronavirus Disease 2019 Case Surveillance—United States, January 22–May 30, 2020. **Morbidity and Mortality Weekly Report**, Centers for Disease Control and Prevention, v. 69, n. 24, p. 759, 2020.

STROGATZ, Steven H. Exploring complex networks. **nature**, Nature Publishing Group, v. 410, n. 6825, p. 268–276, 2001.

STURNIOLO, Simone; WAITES, William; COLBOURN, Tim; MANHEIM, David; PANOVSKA-GRIFFITHS, Jasmina. Testing, tracing and isolation in compartmental models. **PLoS Computational Biology**, Public Library of Science, v. 17, 3 Mar. 2021. `https://doi.org/10.1371/journal.pcbi.1008633`. ISSN 15537358.

TARDÁGUILA, Cristina; BENEVENUTO, Fabricio; ORTELLADO, Pablo. Fake News Is Poisoning Brazilian Politics. WhatsApp Can Stop It. **The New York Times**, v. 17, n. 10, 2018.

THALER, R.H.; SUNSTEIN, C.R. **Nudge: Improving Decisions About Health, Wealth, and Happiness**. [S.l.]: Penguin Publishing Group, 2009. ISBN 9781101655092.

THE MATHWORKS, Inc. **Fuzzy c-means clustering - MATLAB fcm**. [S.l.: s.n.], 2022. Available at: <`https://www.mathworks.com/help/fuzzy/fcm.html`>. Accessed on: 27 Jul. 2022.

TOSIN, Michel; CUNHA JR, Americo; COELHO, Flávio C. Seleção de modelos epidemiológicos via análise de sensibilidade global. **Anais da Sociedade Brasileira de Automática**, v. 2, n. 1, 2020. `https://doi.org/10.48011/asba.v2i1.982`.

TURCULEŢ, Mircea. Ethical Issues Concerning Online Social Networks. **Procedia - Social and Behavioral Sciences**, v. 149, p. 967–972, 2014. LUMEN 2014 - From Theory to Inquiry in Social Sciences, Iasi, Romania, 10-12 April 2014. ISSN 1877-0428.

TUTTE, W. T. **Connectivity in Graphs**. 1. ed. London: Oxford University Press; University of Toronto Press, 1966. v. 15, p. 145. (Mathematical Expositions).

VALLOR, Shannon. Social Networking and Ethics. In: ZALTA, Edward N. (Ed.). **The Stanford Encyclopedia of Philosophy**. Winter 2016. [S.l.]: Metaphysics Research Lab, Stanford University, 2016.

VAN DIJCK, José. **The culture of connectivity: A critical history of social media**. [S.l.]: Oxford University Press, 2013.

VAN MIEGHEM, Piet. The N-intertwined SIS epidemic network model. **Computing**, Springer, v. 93, n. 2-4, p. 147–169, 2011.

VASCONCELOS, Giovani L; DUARTE-FILHO, Gerson C; BRUM, Arthur A; OSPINA, Raydonal; ALMEIDA, Francisco AG; MACÊDO, Antônio. Situation of COVID-19 in Brazil in August 2020: An Analysis via Growth Models as Implemented in the ModInterv System for Monitoring the Pandemic. **Journal of Control, Automation and Electrical Systems**, Springer, p. 1–19, 2022. `https://doi.org/10.1007/s40313-021-00853-3`.

VIGUERIE, Alex; LORENZO, Guillermo; AURICCHIO, Ferdinando; BAROLI, Davide; HUGHES, Thomas J.R.; PATTON, Alessia; REALI, Alessandro; YANKEELOV, Thomas E.; VENEZIANI, Alessandro. Simulating the spread of COVID-19 via a spatially-resolved susceptible–exposed–infected–recovered–deceased (SEIRD) model with heterogeneous diffusion. **Applied Mathematics Letters**, Pergamon, v. 111, p. 106617, Jan. 2021. `https://doi.org/10.1016/J.AML.2020.106617`. ISSN 0893-9659.

VON LUXBURG, Ulrike. A tutorial on spectral clustering. **Statistics and computing**, Springer, v. 17, n. 4, p. 395–416, 2007.

VOSOUGHI, Soroush; ROY, Deb; ARAL, Sinan. The spread of true and false news online. **Science**, American Association for the Advancement of Science, v. 359, n. 6380, p. 1146–1151, 2018.

WANG, Patrick; ANGARITA, Rafael; RENNA, Ilaria. Is this the era of misinformation yet: combining social bots and fake news to deceive the masses. In: INTERNATIONAL WORLD WIDE WEB CONFERENCES STEERING COMMITTEE. COMPANION Proceedings of the The Web Conference 2018. [S.l.: s.n.], 2018. P. 1557–1561.

WATSON, Tom. 'Fake news' is changing the way we see the world. We have to face up to what that means. **Independent UK**, 2016. Available at: <`https://www.independent.co.uk/voices/fake-news-facebook-twitter-social-media-sharing-changing-way-see-world-face-it-labour-tom-watson-a7431466.html`>. Accessed on: 9 Oct. 2019.

WATTS, Duncan J. **Six Degrees: The Science of a Connected Age**. [S.l.]: W. W. Norton, 2004. ISBN 9780393076127.

WATTS, Duncan J.; STROGATZ, Steven H. Collective dynamics of 'small-world' networks. **Nature**, Nature Publishing Group, v. 393, n. 6684, p. 440–442, 1998.

WILSON, Robin J. **Introduction to Graph Theory**. 5. ed. Essex: Pearson, 2010. P. 184. ISBN 978-0-273-72889-4.

WISE, Jacqui. COVID-19: New coronavirus variant is identified in UK. **British Medical Journal**, BMJ Publishing Group Ltd, v. 371, 2020. `https://doi.org/10.1136/bmj.m4857`.

WORK, Dan; WIT, Carlos Canudas de; SINOPOLI, Bruno. Resilient Cyber-Physical Systems. In: ANNASWAMY, Anuradha M.; JOHANSSON, Karl H.; PAPPAS, George J. (Eds.). **Control for Societal-Scale Challenges: Road Map 2030**. [S.l.]: IEEE Control Systems Society, 2023.

WU, Joseph T; LEUNG, Kathy; LEUNG, Gabriel M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. **The Lancet**, Elsevier, v. 395, n. 10225, p. 689–697, 2020. `https://doi.org/10.1016/S0140-6736(20)30260-9`.

ZANKER, Markus; ROOK, Laurens; JANNACH, Dietmar. Measuring the impact of online personalisation: Past, present and future. **International Journal of Human-Computer Studies**, Elsevier, v. 131, p. 160–168, 2019.

ZHANG, Shihua; WANG, Rui-Sheng; ZHANG, Xiang-Sun. Identification of overlapping community structure in complex networks using fuzzy c-means clustering. **Physica A: Statistical Mechanics and its Applications**, Elsevier, v. 374, n. 1, p. 483–490, 2007.

ZHANG, Xichen; GHORBANI, Ali A. An overview of online fake news: Characterization, detection, and discussion. **Information Processing & Management**, p. 102025, 2019. ISSN 0306-4573.