Modelling group dynamics in epidemic opinion propagation

Dieter Fiems

Ghent University, Department of Telecommunications and Information Processing
St-Pietersnieuwstraat 41, 9000 Gent, Belgium
http://telin.ugent.be

Abstract. Motivated by weblogs and discussion forums, epidemic opinion propagation on affiliation networks is investigated. An affiliation network is a bi-partite graph describing the connections between individuals and their affiliations. In contrast to epidemics on complex networks, the epidemic spreading process in the current setting is not the consequence of pairwise interactions among individuals but of a group dynamic. We derive a Markov model for the epidemic process and its fluid limit obtained by sending the population size to infinity while keeping the number of affiliations constant. This results in a set of modified SIR-like ordinary differential equations. Different types of group dynamics are studied numerically and the accuracy of the fluid limit is verified by simulation.

1 Introduction

With the emergence of social network services (SNS), the speed and outreach of information diffusion has reached unprecedented heights. In just over a decade, SNS's have attracted millions of users, many of them using these services on a daily basis [1]. A typical SNS allows users to create a profile and make connections to other users in the social network. A profile is a unique page where one can "type oneself into being" [2] and can be public or semi-public. SNS users can send private messages to their connections, inform their connections when their profile is updated, or pass on messages received from their connections, etc. Such functionality greatly facilitates quick dissemination of information.

This paper studies epidemic-like opinion propagation on social networks. While initial epidemiological models assumed well-mixed populations, it has been in-

creasingly recognised that topological properties of the network of members of the population and their connections greatly affect the epidemic spreading process [3]. The interplay between topology and dynamics is one of the most pressing challenges in the development of network science [4] and runs in parallel with the increased research effort on complex networks [5]. Indeed, the large amount of scientific effort devoted to this subject [6, 7] has made it evident that dynamical processes (like epidemics) taking place on top of a complex network can be strongly influenced by the topological features of the network, especially in the case of scale-free networks, in which the degree distribution (the degree of a node is the number of nodes it is connected to) follows a power law [8].

In contrast to previous studies on epidemic processes on complex networks, we adopt the affiliation network (AN) paradigm [9], which was studied for SNSs in [10] and [11]. An AN describes the connections between individuals and their affiliations. An affiliation can be a shared interest or personal affinity, a common collective activity, etc. [12]. The AN is a bi-partite graph of individuals and affiliations. Such a graph consists of affiliations and individuals and only interconnections between individuals and affiliations are allowed. As opposed to standard complex networks, ANs allow for a considerably richer and a more intricate interaction between individuals. Whereas interaction between individuals is explicitly pairwise in complex networks, multiple individuals can interact jointly by sharing an affiliation in an AN.

The effects of such non-pairwise interaction is the subject of the present study. Borrowing from epidemiological terminology, it is assumed that the state of any individual is either susceptible (S), infected (I) or recovered (R). Such epidemiological models are usually referred to as SIR-type models. The SIR model assumes that an individual's state goes from susceptible to infected to recovered, an infection being the consequence from contact with infected individuals. This process can be directly reformulated in terms of the propagation of opinions on a particular topic: a susceptible individual has yet to form an opinion on a certain topic, whereas infected or opinionated individuals do have such an opinion and spread their opinion to other individuals. Finally, individuals loose their interest in the topic after some time and stop spreading their opinion, which corresponds to recovery in the epidemiological context [13]. While we retain the classical assumption of Markovian SIR models that individuals recover after an exponentially

distributed amount of time, we modify the infection process as to reflect "group dynamics" associated with affiliations. We adopt the term "group dynamics" as introduced by Lewin [14] as the spreading process is not simply the result of the sum of individual interactions [15]. In particular, we assume that affiliations infect their members with a rate which is a generic function of the states of the affiliation's members. That is, if an affiliation has x_S susceptible and x_I infected members, the affiliation's susceptible members get infected with rate $\alpha(x_S, x_I)$, α being a generic function. Obviously, an individual can have multiple affiliations, and it is assumed that infection by the different affiliations are independent processes, such that the infection rate of an individual is the sum of the infection rates of this individual's affiliations.

The remainder of this paper is organised as follows. The epidemic Markov model and the notational conventions of the paper are introduced in the next section. The fluid limit of the Markov model, which is obtained by increasing the population size while keeping the number of affiliations constant, is discussed in section 3 and numerically investigated in section 4. Finally, conclusions are drawn in section 5.

2 Markovian epidemic model

We consider epidemic opinion propagation on ANs. An AN is a bipartite graph, whose vertices are divided into affiliations and individuals and whose edges connect affiliations with individuals.

Let $\mathcal{A}$ be the set of all affiliations and let $\widehat{\mathcal{G}} = \mathcal{P}(\mathcal{A})$ be the power set of $\mathcal{A}$, that is $\widehat{\mathcal{G}}$ is the set of all subsets of $\mathcal{A}$. Further, let $\mathcal{X}$ be the set of individuals. Each individual can have multiple affiliations, for an individual $i \in \mathcal{X}$, let $g(i) : \mathcal{X} \to \widehat{\mathcal{G}}$ be the set of this individual's affiliations. The mapping g induces a partition of $\mathcal{X}$, all individuals having the same affiliations in each subset of the partition. For $G \in \widehat{\mathcal{G}}$, let $\mathcal{X}_G = \{x \in \mathcal{X}, g(x) = G\}$ be the corresponding subset of $\mathcal{X}$ and let $N_G = |\mathcal{X}_G|$ be the number of individuals in this subset. For any set X, $|X|$ denotes its cardinality. We may exclude subsets G with $N_G = 0$ from further analysis. Therefore, let $\mathcal{G} = \{G \in \widehat{\mathcal{G}} : N_G > 0\}$.

With a slight abuse of notation, for any affiliation $a \in \mathcal{A}$, let $\mathcal{X}_a$ be the set of individuals having affiliation a, $\mathcal{X}_a = \{x \in \mathcal{X} : a \in g(x)\}$, and let $N_a = |\mathcal{X}_a|$ be the

number of individuals in this set. Note that for $a_1 \neq a_2$ the intersection of $\mathcal{X}_{a_1}$ and $\mathcal{X}_{a_2}$ may be non-empty as individuals may have affiliations a_1 and a_2.

We adopt a Markovian SIR-type epidemic process. At any time, an individual is in one out of three possible states: susceptible, infected or recovered. Hence, the individuals can also be partitioned into susceptible, infected and recovered individuals. Let $\mathcal{S}(t), \mathcal{I}(t)$ and $\mathcal{R}(t)$ be the sets of susceptible, infected and recovered individuals at time t, and let

$$S_G(t) = |\mathcal{S}(t) \cup \mathcal{X}_G| \,, \quad I_G(t) = |\mathcal{I}(t) \cup \mathcal{X}_G| \,, \quad R_G(t) = |\mathcal{R}(t) \cup \mathcal{X}_G| \,.$$

Individuals in the same partition $G \in \mathcal{G}$ are indiscernible. Moreover, affiliations inherit their state from the state of their members. Therefore, the state of the epidemic process is completely described by the number of susceptible and infected individuals in the different subsets $G \in \mathcal{G}$. Let $\mathbf{S}(t) = [S_G(t)]_{G\in\mathcal{G}}$ and $\mathbf{I}(t) = [I_G(t)]_{G\in\mathcal{G}}$ be the vectors whose elements represent the number of susceptible and infected individuals in the different partitions at time t. Here and in the remainder, we index vectors by the elements of $\mathcal{G}$ for ease of presentation. Moreover, let $\pi(\mathbf{s}, \mathbf{i}; t) = \Pr[\mathbf{S}(t) = \mathbf{s}, \mathbf{I}(t) = \mathbf{i}]$, for $\mathbf{s} = [s_G]_{G\in\mathcal{G}}$ and $\mathbf{i} = [i_G]_{G\in\mathcal{G}}$, such that $(\mathbf{s}, \mathbf{i}) \in \mathcal{N}$. Here $\mathcal{N}$ denotes the state space of the Markov chain,

$$\mathcal{N} = \{([s_G]_{G\in\mathcal{G}}, [i_G]_{G\in\mathcal{G}}) : s_G, i_G \in \mathbb{N}, s_G + i_G \leq N_G\} \,.$$

For $a \in \mathcal{A}$ and given state vectors $\mathbf{s}$ and $\mathbf{i}$, let $s_a(\mathbf{s})$ and $i_a(\mathbf{i})$ be the fraction of susceptible and infected individuals that have affiliation a,

$$i_a(\mathbf{i}) = \frac{1}{N_a} \sum_{G\in\mathcal{G}, a\in G} i_G \,, \quad s_a(\mathbf{s}) = \frac{1}{N_a} \sum_{G\in\mathcal{G}, a\in G} s_G \,.$$

Affiliation $a \in \mathcal{A}$ infects its susceptible members with a rate $\alpha_a(s_a(\mathbf{s}), i_a(\mathbf{i}))$, α_a being a generic function. The infection rate experienced by individuals in the subset $G \in \mathcal{G}$ therefore equals,

$$\beta_G(\mathbf{s}, \mathbf{i}) = \sum_{a\in G} \alpha_a(s_a(\mathbf{s}), i_a(\mathbf{i})) \,.$$

Let γ be the recovery rate of the individuals, the Chapman-Kolmogorov equations are then given by,

$$\begin{aligned} \frac{d}{dt}\pi(\mathbf{s}, \mathbf{i}; t) = &\sum_{G\in\mathcal{G}} \pi(\mathbf{s} + \mathbf{e}_G, \mathbf{i} - \mathbf{e}_G; t)\beta_G(\mathbf{s} + \mathbf{e}_G, \mathbf{i} - \mathbf{e}_G)(s_G + 1) \\ &+ \sum_{G\in\mathcal{G}} \pi(\mathbf{s}, \mathbf{i} + \mathbf{e}_G; t)\gamma(i_G + 1) - \pi(\mathbf{s}, \mathbf{i}; t) \sum_{G\in\mathcal{G}} (\gamma i_G + \beta_G(\mathbf{s}, \mathbf{i})s_G) \,, \end{aligned}$$

where we set $\pi(\mathbf{s}, \mathbf{i}; t) = 0$ for $(\mathbf{s}, \mathbf{i}) \notin \mathcal{N}$ to simplify notation. Moreover $\mathbf{e}_G$ is a vector of zeros apart from the Gth element which equals 1. The first term on the right-hand side of the former expression corresponds to an infection of an individual in one of the sets $G \in \mathcal{G}$. The second term corresponds to having a recovery in these different sets.

3 Fluid limit

Due to the considerable size of the state space $\mathcal{N}$, even for modest population sizes and a modest number of affiliations, direct computation of either transient or stationary distributions is quite forbidding. As we are mainly interested in the dynamics when the population is large, we focus on the fluid limit of the process. The present study scales the size of the population, while keeping the number of affiliations constant. Let $\mathcal{F}$ be the infinitesimal generator of the Markov process above, we then have,

$$\mathcal{F}h(\mathbf{s}, \mathbf{i}) = \sum_{G \in \mathcal{G}} [h(\mathbf{s} - \mathbf{e}_G, \mathbf{i} + \mathbf{e}_G) - h(\mathbf{s}, \mathbf{i})] \beta_G(\mathbf{s}, \mathbf{i}) s_G + [h(\mathbf{s}, \mathbf{i} - \mathbf{e}_G) - h(\mathbf{s}, \mathbf{i})] \gamma i_G \,.$$

We now consider a sequence of Markov chains with generators $\mathcal{F}_N$ such that the number of individuals is N for the Nth Markov chain, thereby equally scaling N_G for the different sets G; set $\nu_G = \lim_{N \to \infty} N_G N^{-1}$. We track the fractions of populations, such that components of the state space $\mathcal{N}_N$ of the Nth Markov chain live on a lattice with step size $1/N$, the unit vectors having size $1/N$ as well. In contrast, the transition rates increase by N as we translate from population fractions to population sizes. Setting $\epsilon \doteq 1/N$, we get the following generator:

$$\begin{aligned} \mathcal{F}_{\epsilon^{-1}} h(\mathbf{s}, \mathbf{i}) = \epsilon^{-1} \sum_{G \in \mathcal{G}} & [h(\mathbf{s} - \epsilon \mathbf{e}_G, \mathbf{i} + \epsilon \mathbf{e}_G) - h(\mathbf{s}, \mathbf{i})] \beta_G(\epsilon^{-1} \mathbf{s}, \epsilon^{-1} \mathbf{i}) s_G \\ + \epsilon^{-1} \sum_{G \in \mathcal{G}} & [h(\mathbf{s}, \mathbf{i} - \epsilon \mathbf{e}_G) - h(\mathbf{s}, \mathbf{i})] \gamma i_G n \,. \end{aligned}$$

We can deduce the (candidate) fluid limit by Taylor expansion of this generator around $\epsilon = 0$. We find a limiting generator of the form $\hat{\mathcal{F}} h = \mathbf{f}(\mathbf{x}, \mathbf{y}) \cdot \nabla h$, for a certain $2|\mathcal{G}|$-dimensional vector function $\mathbf{f} = [\mathbf{f}_1, \mathbf{f}_2]$. Note that a generator of this form corresponds to a deterministic process satisfying the system of differential equations $\dot{\mathbf{x}}(t) = \mathbf{f}_1(\mathbf{x}(t), \mathbf{y}(t))$, $\dot{\mathbf{y}}(t) = \mathbf{f}_2(\mathbf{x}(t), \mathbf{y}(t))$.

In order to prove this limit rigorously, it needs to be checked that both the pre-limit processes and the limit process are Feller processes [16], which corresponds

to checking the Hille-Yosida conditions. We believe that a careful proof falls outside the scope of this paper, but remark that due to the compactness of the state space the proof is not as involved as is sometimes the case. Below we detail the set of differential equations, where we have dropped the dependence on t for notational convenience. For all $G \in \mathcal{G}$, we have,

$$s'_G = -\widehat{\beta}_G(\mathbf{i}, \mathbf{s}) s_G \,, \quad i'_G = \widehat{\beta}_G(\mathbf{i}, \mathbf{s}) s_G - \gamma i_G \,, \quad r'_G = \gamma i_G \,,$$

where s_G, i_G and r_G are the fraction of susceptible, infected and recovered individuals that have affiliation set G, respectively. Here $\widehat{\beta}$ couples the differential equations for the different affiliation sets as follows,

$$\widehat{\beta}_G(\mathbf{i}, \mathbf{s}) = \sum_{a \in G} \alpha_a \left(\frac{1}{\nu_a} \sum_{H \in \mathcal{G}, a \in H} i_H, \frac{1}{\nu_a} \sum_{H \in \mathcal{G}, a \in H} s_H \right) ,$$

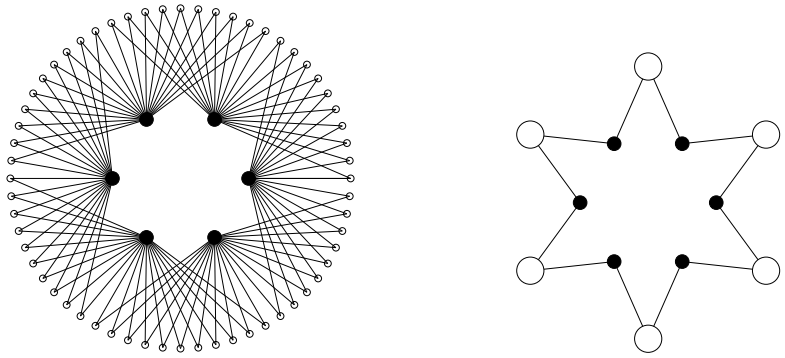
with $\nu_a = \lim_{N \to \infty} N_a N^{-1}$.

4 Numerical examples

We adopt the topology of Fig. 1 for the numerical examples. The affiliations and individuals live on circles, and an individual connects to its κ closest affiliations, the distance being measured in terms of difference in angle between individual and affiliation. In addition, we assume the same group dynamic in each affiliation and the infection rate of the affiliations only depends on the fraction of infected in the affiliation.

We focus on *regular* dynamics, in which case the infection rate is an increasing function of the number of infected, as well as on *early adopter dynamics* in which case the infection rate decreases if more members of the affiliation are infected.

Figure 2 assesses the accuracy of the fluid approximation by means of simulation. All plots depict the time-evolution of the percentage of susceptible (S), infected (I) and recovered (R) individuals in the population. The lines correspond to the fluid limit, whereas the markers correspond to a single trajectory of the epidemics, obtained by simulating the Markov chain. The population size is $N = 1000$ in figures 2(a) and 2(b), $N = 10000$ in 2(c) and 2(d), and $N = 100000$ in 2(e) and 2(f). All individuals have 3 affiliations, thereby assuming the topology of figure 1. The initial infection consists of 1% of infected individuals that share the same affiliations. All other individuals are susceptible. The infection rate function is regular



(a) individual connections (b) group connections

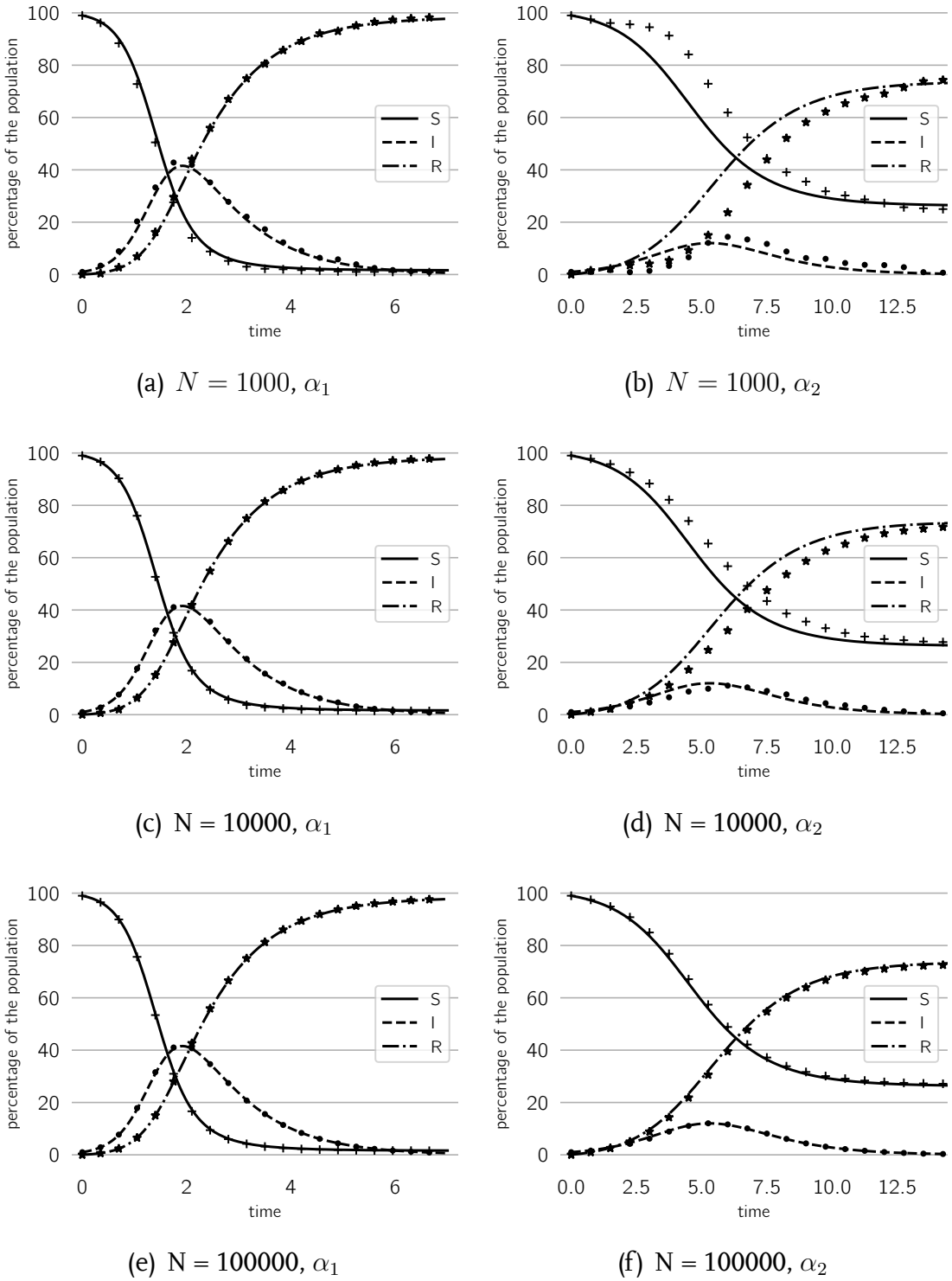
Fig. 1. Circular structure for a network with 60 individuals (outer circle) and 6 affiliations (inner circle), each individual having two affiliations. Figure (a) shows the individual connections, figure (b) groups the individuals with the same affiliations.

and superlinear in Figs. 2(a), 2(c) and 2(e), $\alpha_1(i) = 1.4i - \mathbb{1}_{\{i>1/2\}}0.8(i-1/2)$, and regular and sublinear in Figs. 2(b), 2(d) and 2(f), $\alpha_2(i) = 0.6i + \mathbb{1}_{\{i>1/2\}}0.8(i-1/2)$. Finally the recovery rate is $\gamma = 1$ for all plots. There is clear discrepancy between the plots with super- and sublinear dynamics, the infection for the superlinear case being considerably more extensive. In either case, the fraction of infected is always less than 50% such that the slope of the infection rate function for $i < 1/2$ entirely determines the dynamics of the epidemic. Simulation confirms the accuracy of the fluid limit for $N = 10^5$.

We now compare regular and early adopter dynamics. Figure 3 depicts the time-evolution of the percentage of susceptible, infected and recovered individuals for regular ($\alpha_3(i) = 2i\nu + \mathbb{1}_{\{i>1/2\}}2(1-2\nu)(i-1/2)$) dynamics and for early adopter dynamics ($\alpha_4(i) = 1 - 2i(1-\nu) + \mathbb{1}_{\{i>1/2\}}2(1-2\nu)(i-1/2)$). Here, ν is the value of α for $i = 0.5$; different values of ν are assumed as indicated. A comparison of the curves of regular dynamics and early adopter dynamics reveals that the speed and the maximal size of the infection for early adopter dynamics is faster and larger than regular dynamics. This is not unexpected as the infection rate is larger at the onset of the infection for early adopter dynamics.

5 Conclusion

We proposed an epidemic process on an affiliation network for modelling group dynamics for opinion propagation on social networks. Opinions are spread from

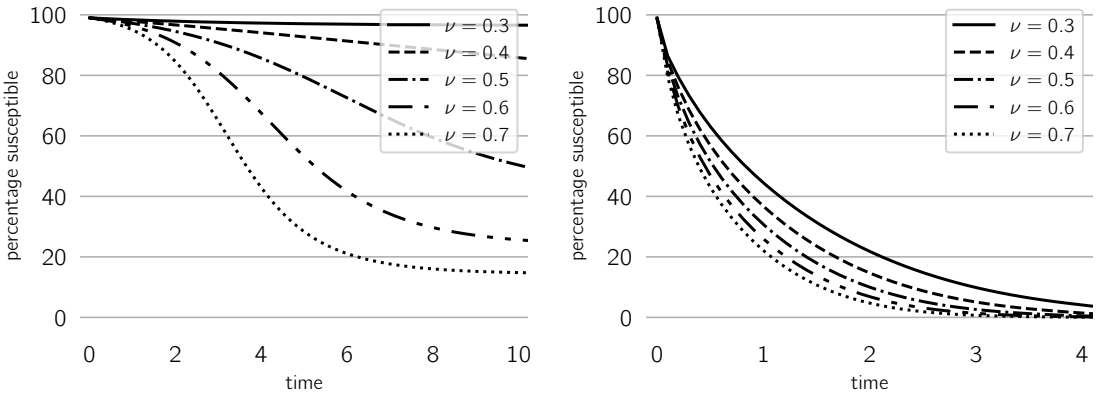


(a) $N = 1000$, α_1 (b) $N = 1000$, α_2

(c) N = 10000, α_1 (d) N = 10000, α_2

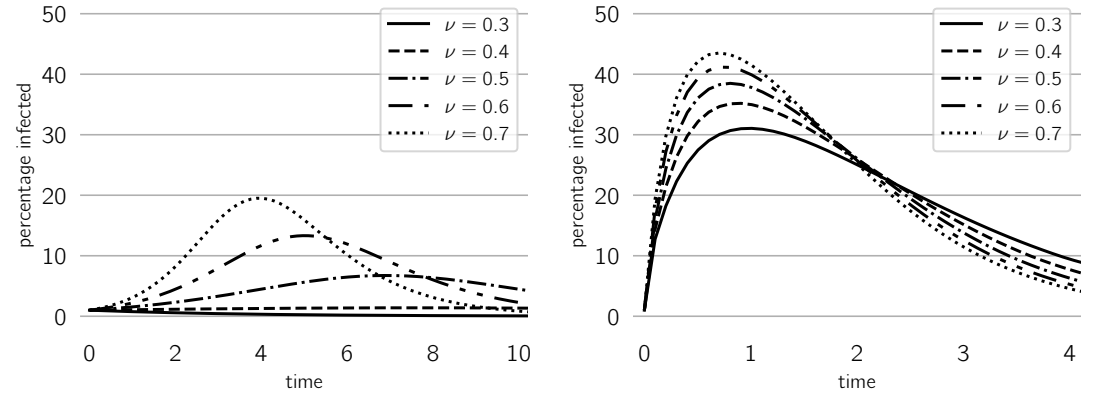
(e) N = 100000, α_1 (f) N = 100000, α_2

Fig. 2. Accuracy of the fluid limit for a sub- and super-linear infection rate function.



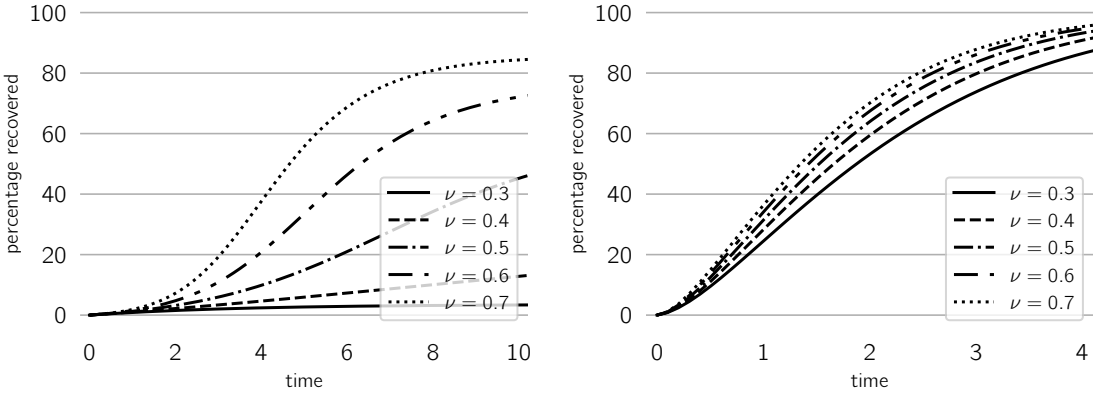
(a) regular, α_3, susceptible

(b) early adopters, α_4, susceptible



(c) regular, α_3, infected

(d) early adopters, α_4, infected



(e) regular, α_3, recovered

(f) early adopters, α_4, recovered

Fig. 3. Regular dynamics versus early adopter dynamics for different values of ν.

one individual to another via shared affiliations: the opinions of the members of an affiliation determine the spread of the opinions to the (non-infected) members of the affiliation. We provided a continuous-time Markov process for SIR-like propagation, and studied its fluid limit. That is, we scaled the Markov process by sending number of individuals to infinity while keeping the number of affiliations constant. By numerical examples, we showed that the fluid limit is accurate when the number of individuals is sufficiently large, while the nature of the group dynamic can seriously affect spreading in the network.

Apart from the SIR epidemic, other epidemic models may apply to rumour spreading as well. For example, if the SIS model is adopted, individuals alternate between being susceptible and infected, i.e. between spreading and not spreading their opinion. In the SEIR model, individuals are exposed before they are infected, which introduces some time during which in individual has adopted the opinion, but does not yet spread. We aim at developing similar mathematical tools for these alternative epidemic processes on affiliation networks in the near future.

References

1. D.M. Boyd and N.B. Ellison. Social network sites: definition, history, and scholarship. *Journal of Computer-Mediated Communication* 13:210-230, 2008.
2. J. Sundén. *Material Virtualities*. New York: Peter Lang. 2003.
3. K. Avrachenkov, K. De Turck, D. Fiems, and B.J. Prabhu. Information dissemination processes in directed social networks. International Workshop on Modeling, Analysis and Management of Social Networks and their Applications (SOCNET). MMB & DFT 2014, Bamberg, Germany, 2014.
4. M.E.J. Newman. *Networks: An introduction*. Oxford University Press, 2010.
5. A.-L. Barabási and E. Bonabeau. Scale-free networks. *Scientific American*, 288:50-59, 2003.
6. S.N. Dorogovtsev, A.V. Goltsev, and J.F.F. Mendes. Critical phenomena in complex networks. *Reviews of Modern Physics* 80: 1275-1335, 2008.
7. A. Barrat, M. Barthélemy, and A. Vespignani. *Dynamical processes on complex networks*. Cambridge University Press, 2008.
8. A.-L. Barabási, R. Albert, H. Jeong, and G. Bianconi . Power-law distribution of the world wide web. *Science* 287:2115a, 2000.
9. S. Lattanzi and D. Sivakumar, Affiliation networks. In: Proceedings of the 41st ACM Symposium on Theory of Computing, pp. 427-434, Maryland, USA, 2009.
10. E. Zheleva, H. Sharara, and L. Getoor. Co-evolution of social and affiliation networks. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1007-1016, Paris, France, 2009.

11. S. Ghosh, S. Saha, S. Srivastava, T. Krueger, N. Ganguly, and A. Mukherjee. Understanding evolution of inter- group relationships using bipartite networks. *IEEE Journal on Selected Areas in Communciations* 31(9):584-594, 2013.
12. R.L. Breiger. Duality of persons and groups. *Social Forces* 53(2):181-190, 1974.
13. E. De Cuypere, K. De Turck, S. Wittevrongel, D. Fiems. Opinion propagation in bounded medium-sized populations. *Performance Evaluation* 99–100:1–15, 2016.
14. K. Lewin. Frontiers in Group Dynamics: Concept, Method and Reality in Social Science; Social Equilibria and Social Change. *Human Relations* 1:5–41, 1947.
15. D.R. Forsyth. *Group Dynamics*. 6th ed. Wadsworth Publishing, 2013.
16. S. N. Ethier and T. G. Kurtz. *Markov processes.* John Wiley & Sons, 1986.