Organisational Structure

Ka Man (Ambrose) Yim

University of Oxford

Supervisors: Renaud Lambiotte, Andrew Mellor,
Stephen Brewis, Stephen Cassidy

A technical report for

InFoMM CDT Mini-Project 1

in partnership with

BT

Trinity 2018

Contents

1 Introduction

We can think of an organisation as a problem solving machinery. A problem facing a company like Ford or BMW might be to build a certain number of cars in a month to a certain specification; the problem facing BBC is to make programmes that fit their organisational objectives to 'inform, educate and entertain', subject to regulatory oversight and budget restrictions. A restaurant's problem is to make delicious food at an attractive price *sans e. coli.*

The type of problems an organisation tackles are those which are too large and complex to be solved by a single person with reasonable efficiency. Hence labour is divided between many people, each person serving some function with different capabilities. This process is analogous to decomposing computations on a computer into several processes and solving them in parallel on multiple processing cores. However, unlike computer processes that execute a fixed set of instructions, humans (allegedly) have free will and agency. Thus an organisation makes its own decisions and has a life of its own through the manifold interactions and emails between people. Decision-making in organisations are rarely made in isolation by a single person. Behind each decision there is a complex web of influence and information transfer (or lack thereof). In most (if not all) levels of an organisation, decisions are reached by some form of consensus.

Should we trust humans to make decisions that best align with the global objectives of the organisation? People often fail to see beyond the costs and benefits arising from their interactions with their *local* environments; few individuals have a holistic understanding of their role in the global picture of interactions. We live in a remarkable time in history (which is decidedly not at its end) when democratic decisions, entangled with advances in communication and new ways of influencing people, have shook the Western liberal-democratic political system to its core.

Nonetheless in the pre-SkyNet era, humans are the only candidates as the atoms of organisations. The question is how people should be organised to make 'better' decisions. Here we present a mathematical model for decision making processes in an organisation. We assume in a corporate setting there is a well-defined 'global' problem, in the sense that there is a measure to compare how well two solutions solve the same problem. A solution is a set of decisions, of which there are many.

Each person, or agent, has a decision to make. It can choose between several actions. Four characteristics fully define an agent:

1. the individual actions available to the agents (capabilities);

2. the bias of agents towards certain actions;

3. how actions of other agents affect an agent's bias; and

4. who the agent talks with.

The first three features give rise to a global network describing the dependencies between agents. The position of the agent in this **dependency network** defines the

influence relations between agents. The **social network** of the organisation emerges from the last feature.

Ideally, the social network of an organisation should be built around the dependency network; people whose actions are highly correlate with each other should communicate and coordinate. However, agents may not be fully aware of the full extent of the dependency network, diminishing the chance of communication between parties that influence one another. Moreover, there is a social cost involved in communication, and agents can only maintain a limited number of agents in its social sphere; thus given a sufficiently complex problem agents inevitably will fail to communicate with all of its neighbours in the dependency network.

The social network on the same level of organisational hierarchy are often structured into closely knit teams with fairly little communications between teams. In this study we study the efficiency of social networks with fairly distinct social structures in solving problems . The particular problem we are investigating is called a coordination game, where an agent only deems it desirable to coordinate if sufficiently many of its neighbours on the dependency network decide to coordinate. The global objective of the organisation is for as many agents to coordinate. This dynamics of persuasion can be mapped onto threshold contagion models on networks, where a disease or 'idea' spreads along a network of people.

We want to measure how different network topologies in the communications and dependency networks influence how

We formally define our mathematical model below.

2   Model

We suppose there are N agents working in an organisation. We follow the standard game theory agent model. Each agent i has a set of actions A_i and a utility function $u_i : A_1 \times \cdots A_N \to \mathbb{R}$ which assigns to each state of the system, which we define to be the collection of all the actions of the agents, a real value which we call the utility of the state to agent i. Given the actions of all the other agents, each agent i will selfishly try to choose an action out its own set of actions A_i to maximise its own utility.

In our model we assume each agent's utility is only biased by a small number of agents in the system. We construct a dependency network of influence and constraints by drawing an edge between agents whose utilities depend on one another. In our simplification we assume that inflence is reflexive i.e. $u_i = u_i(a_i, a_j, \ldots) \Leftrightarrow u_j = u_j(a_j, a_i, \ldots)$, such that the dependency network is undirected.

As a simple model for agents adapting to new ideas in an organisation we consider the coordination game [8] [4] being played out on a dependency network. In the coordination game each agent has two actions $A_i = \{0, 1\}$, where $a_i = 0$ is what we call the default action and $a_i = 1$ is the novel action, which represents innovation. If the agent chooses the default action $a_i = 0$, the utility will always be neutral $u_i(a_i = 0 \mid a_{-i}) = 0$. However, if the agent elects to implement the novel action, two scenarios may arise. If a sufficiently large fraction $r_i \in [0, 1]$ of its neighbours in

the dependency network decide to coordinate, then the agent will receive a positive utility $u_i > 0$, and the logical action of the agent would then be to commit to the novel action. However if there are fewer than a fraction r_i of neighbours in the neighbourhood to prompt the agent to innovate, the $u_i < 0$, and the best response of the agent would be return to the default action. This simple dynamic simulates the resistance to the uptake of new ideas, and the effects of constraints imposed by other agents in an organisation in deterring innnovation. We say the global objective of the organisation is to have as many agents adopt the novel action.

To characterise a population's tendency to accept new ideas, we suppose each agent samples their persuasion threshold r_i from some distribution. Following Gleeson's example [2], we suppose the distribution is a Gaussian with standard deviation $\sigma = 0.2$ and mean R, a parameter which we tune to adjust the characteristic of the population. Agents with $r_i < 0$ will always elect to adopt new ideas. Gleeson [2] aptly refers to these agents as innovators. We need innovators to initiate the spread of $a = 1$ across the population of agents. As we increase R, there will be fewer innovators in the population and the aggregate tolerance of new ideas is decreased.

In the case where the social network matches the dependency network, the only condition for innovation to 'infect' an agent is for the agent to be exposed to sufficiently many innovative neighbours is the dependency network. However, once we decouple the social network from the dependency network, the agents need to prompted by neighbours in its social network to be aware of innovation and consider adopting it. Thus an agent is in one of three states at any time: unaware and susceptible, aware and susceptible, and aware and infected. In this regard our model is similar to the multiplex SIR contact infection models in [3]. We study two types of social interactions. In the first case, the social awareness threshold of a node, which is the smallest fraction of innovating neighbours to make a previously unaware node aware of innovation, is sampled independently from the same distribution as the depedendency threshold. This is the model studied by Lee, Brummitt and Goh [6]. In the second case, an agent is aware if only one of its neighbours decides to innovate. The model we study here is a special case of the

We consider two ways which the social and dependency networks could decouple. We could take a social network and either (a) swap pairs of nodes, or (b) swap pairs of edges to obtain the dependency network (or vice versa). (a) represents the social network failing to catch up with changes in role of agents the dependency network, while (b) represents an actual change in dependency. In (a), the perturbed dependency network has the same network topology but a different correspondence between nodes between the social and dependency layers, whereas in (b) we can think of the nodes maintaing the same correspondence between layers but the network topology being altered as the two networks decouple.

The nodes update their actions according to the following rules which is widely accepted as a standard in network epidemiology literature [9] [2] and network game theory literature [4]. Firstly, the initial infected population is identified: these may be 'seeds' of coordinating agents which we put in by hand, or innovators with negative thresholds. Then nodes in the 0 state update their responses synchronously according to the information they receive and the agents whom they depend on. Once the

nodes decide to innovate, they will remain in the state of innovation. We use the final fraction of infected agents ρ as a measure of the success of the spread of innovation in the organisation. In our simulations we will see how ρ decreases as the number of node or edge swaps increase in both (a) and (b) modes of decoupling. In our analytical study we compute the 'cascade threshold', the maximal value of R greater than which the population would not adopt new behaviour in the limit of infinite swaps. We compare the single layer case to the duplex social + dependency network model, and observe the influence of decoupled social awareness on the cascade of innovation.

We consider two classes of random networks. The first class are Erdős-Renyi graphs [1] where an edge between two nodes is present in the graph with probability p. The second class of networks are 'planted partition models' [5], where agents are partitioned into tightly connected communities with only a few edges between agents int different communities. The first class is amenable to analytical analysis while we will make empirical observations about the properties of the latter from simulations.

3 Theory Review

3.1 Game-theoretic results

3.2 Watts-Gleeson theory of single network contagion

The theory of cascades on large random graphs generated by degree distributions following the aforementioned dynamical rules was first studied by Watts [9] and a full account was later given by Gleeson [2]. Their theory rigorously holds for large random graphs that are locally treelike, where clustering (the probability that a triple of connected nodes froms a triangle) is negligible.

Here we summarise the motivations and results of Gleeson's theory [2]. Let us consider a locally tree-like infinite random graph with a degree distribution $p(k)$ and select a node which we designate as the top of the tree. We consider the probability that an infection initiated infinitely far away from the top node at the leaves of a tree infects the top node. Since the node was chosen arbitrarily, this probability is the probability of any node being infected by the cascade initiated by the initial infected population in an infinite network. This probability is also the steady state fraction of infected nodes ρ. We suppose a fraction ρ_0 of nodes are initially infected.

We assume that the nodes update their actions starting from the leaves of the tree, which we designate to be level $n = 0$, and sequentially up the levels from n to $n + 1$ until it reaches the nodes in the neighbourhood of the top node, which we designate to be level $n = \infty$. Let q_n be the probability that a node on level n is infected in the $a = 1$ state conditioned on its parent on level $n + 1$ being in the default $a = 0$ state. Since the nodes are updated sequentially up the levels of the tree, q_{n+1} is only dependent on q_n, and the dependency is mediated by a propagator g

$$q_{n+1} = G(q_n) = \rho_0 + (1 - \rho_0)g(q_n) \quad \text{for} \quad n = 0, 1, 2, \ldots \; ; \; q_0 = \rho_0, \tag{1}$$

6

which is a function dependent on the degree distribution of the network and the threshold conditions but independent of the level of the tree n. Since q is a probability, it is bounded between $q \in [0, 1]$. Therefore $G : [0, 1] \to [0, 1]$ is a map from a unit interval to itself. If G is a continuous function (which it is), then by Brouwer's fixed point theorem there must be at least one fixed point where $q = G(q)$. As the updating process reaches the top of the tree $n \to \infty$ -analogous to evolving a dynamical system to infinite time - q_n must converge on a stable fixed point q_∞. Thus the probability of infecting the top node, or the final fraction of infected nodes ρ, is simply a function of q_∞ and the initially infected population.

We proceed with the derivation of the analytic forms of the functions g and ρ. For notational convenience, we let the binomial probability to be

$$B_m^k(q) = \binom{k}{m} q^m (1 - q^m)^{k-m} \tag{2}$$

and define the response function $F(x)$ to be the probability that the infection threshold of a node is smaller than x.

We start with ρ and suppose the top node in consideration is not in the initial infected population. Then the probability that the node is infected is simply the probability that the infection threshold is less than m/k, where k is the degree of the node and m is the number of infected neighbours. Since the probability of a neighbour being infected is simply q_∞, we write down a binomial distribution for m. We sum over all m and all k to find the probability that an uninfected top node is infected upon update. Taking into consideration that it may be initially infected, we can write down the final fraction as.

$$\rho = \rho_0 + (1 - \rho_0) \sum_{k=1}^{\infty} p_k \sum_{m=0}^{k} B_m^k(q_\infty) F(\frac{m}{k}). \tag{3}$$

$g(q_n)$ is the probability that an uninfected node in layer $n + 1$ is infected when it updates its action. Thus $g(q)$ is much similar to $\rho(q_\infty)$ in form. However since it is given that the node is only infected by its children in layer n, the binomial distribution now runs from $m = 0, \ldots, k - 1$, and the probability distribution for the degree of the node is replaced by the probability of reaching a node of degree k by following an edge from the parent node on level $n + 2$, which is kp_k/z [7], where $z = \langle k \rangle$. Thus the expression for $g(q)$ is

$$g(q) = \sum_{k=1}^{\infty} \frac{k}{z} p_k \sum_{m=0}^{k-1} B_m^{k-1}(q) F(m/k). \tag{4}$$

We are interested in the dynamics in the $\rho_0 \to 0$ limit and thus focus our attention on the dynamical behaviour of the recursion $q_{n+1} \leftarrow g(q_n)$. In particular, we would like to find the fixed points of the recursion and assess its stability. Gleeson demonstrated that in certain regimes of parameter space (z, r), there are multiple fixed points $q_\infty(z, r)$ and these fixed points display bifurcating behaviour. Here we reproduce
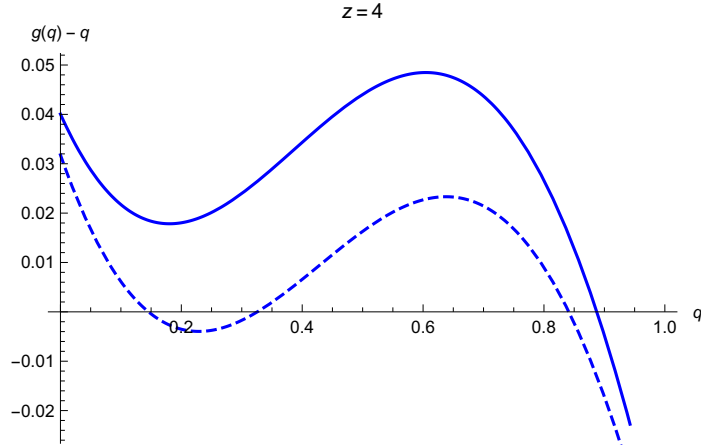
Figure 1: $g(q) - q$ for parameters $z = 4$; sold is $r = 0.35$, dashed is $r = 0.371$

several plots from Gleeson's paper. Let us consider several cross sections of the $q_\infty(z, r)$, where we fix the adoption threshold parameter r and vary the mean degree of the network z. At small values of r, we see a single stable fixed point, with q_∞ smoothly increasing as we increase z. The growth of q_∞ becomes rather steep in some region of z and levels off to a value close to 1 at large values of z. At a critical value of r multiple fixed points appear with alternating stability and we observe several saddle point bifurcations. Since we iterate q from $q_0 \ll 1$, the system will always end up on the lowest possible branch. We observe that in some intermediate regime of the threshold parameter there are discontinuous upward and downward leaps between lower branches of q_∞, where the infection is contained, and the upper branches, where the infection is extensive in the system. In the last plot where r is large, iterating q from 0 will alway lead to system to the lower branch of q_∞ and the infection will never be extensive.

These plots allow us to make a succinct qualitative summary of the effects of network topology and agent thresholds on the spread of ideas in a population of individuals. In the low threshold regime, the increase in network connectivity aids the spread of ideas, while in the intermediate threshold regime there is a 'goldilocks zone' of values of z where ideas can cascade extensively across the population; on either side of agents are either too sparsely connected for the contagion to become extensive or are become overly constrained by the influence of its neighbours too spark an initial adoption. Finally, if the threshold parameter is too large, then the infection will never cascade extensively because agents in the neighbourhood of the few available innovators are too constrained by other neighbours to adopt new ideas. Fixing z across the plots, we observe that as r is increased, a fixed point may drop discontinuously from the upper branch to the lower branch at some critical value of r.

In the duplex system, we follow the extension on Gleeson's work by Brummitt. We have a joint distribution for the degrees in both layers $p(\mathbf{k}) = p(k, k')$, where k denotes the degree in the social network and k' denotes the degree in the dependency network. In the case where the nodes in the social network are swapped to produce
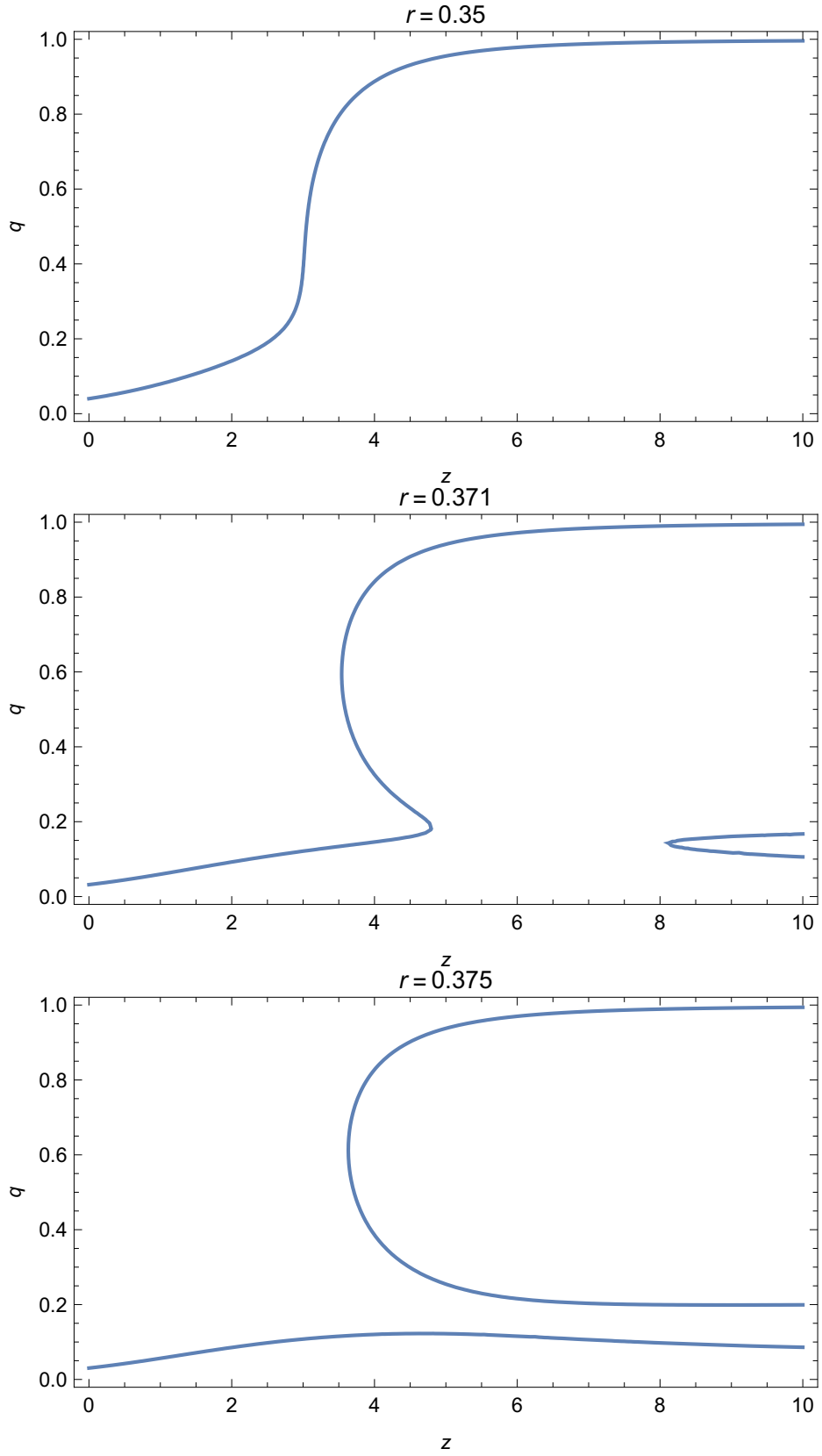
Figure 2: Fixed point plot for increasing r. Note Goldilocks zone in middle figure.

the dependency network $p(\mathbf{k}) \to p(k)p(k')$ in the limit of a large number of swaps; by effectively relabelling the nodes in the network, given a sufficiently large number of nodes, the degree of a node in the dependency network would be uncorrelated with it's degree in social network. In the case where the edges are swapped, the degree in the dependency network is preserved, hence $p(\mathbf{k}) \to p(k)\delta_{k,k'}$ in the large number of swap limit.

Let $F(\mathbf{m}, \mathbf{k})$ be the response function. To simplify the analysis, we assume the threshold for the transfer of awareness in the social network and the threshold for coordination in the dependency network are independently drawn from $\mathcal{N}(r, 0.2)$; hence the response function, which is the probability that the thresholds of the node is below m/k and m'/k' in both networks, is simply

$$F(\mathbf{m}, \mathbf{k}) = \Phi(m/k)\Phi(m'/k') \tag{5}$$

where Φ is the cumulative distribution function for the normal distribution:

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-r)^2}{2\sigma^2}\right). \tag{6}$$

In the duplex situation, we have the expression for final infection fraction

$$\rho = \rho_0 + (1 - \rho_0) \sum_{\mathbf{k}=0}^{\infty} p(\mathbf{k}) \sum_{m=0}^{k} \sum_{m=0}^{k'} B_m^k(q_\infty) B_{m'}^{k'}(q'_\infty) F(\mathbf{m}, \mathbf{k}) \tag{7}$$

where $\mathbf{q}_\infty = (q_\infty, q'_\infty)$ is the fixed point of the recursive iteration

$$\mathbf{q}_{n+1} = \mathbf{g}(\mathbf{q}_n). \tag{8}$$

and the function $\mathbf{g}(\mathbf{q}) = (g(\mathbf{q}), g'(\mathbf{q}))$ is defined to be

$$g(\mathbf{q}) = \rho_0 + (1 - \rho_0) \sum_{k=1}^{\infty} \sum_{k'=0}^{\infty} \frac{k}{z} p(\mathbf{k}) \sum_{m=0}^{k-1} \sum_{m'=0}^{k'} B_m^{k-1}(q) B_{m'}^{k'}(q') F(\mathbf{m}, \mathbf{k}) \tag{9}$$
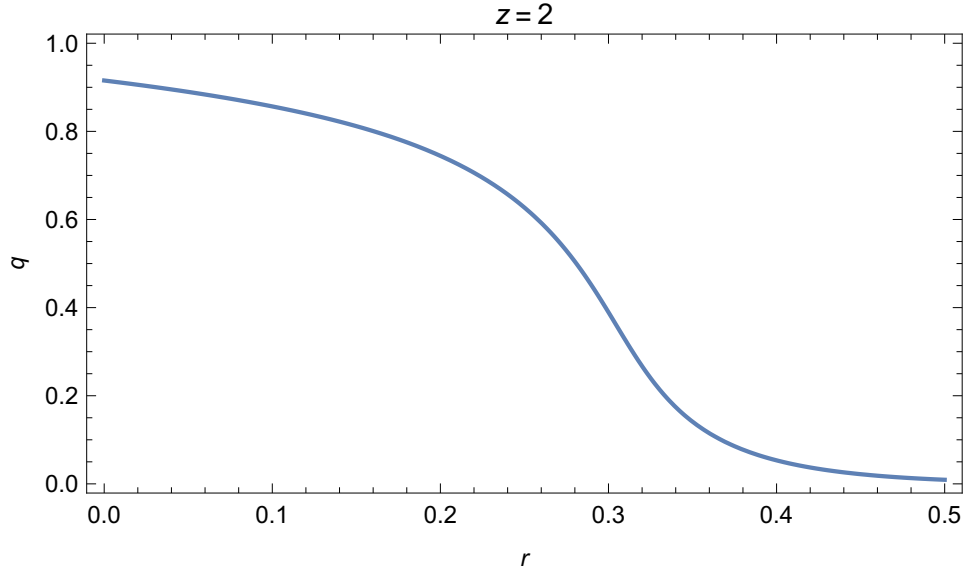
$$g'(\mathbf{q}) = \rho_0 + (1 - \rho_0) \sum_{k'=1}^{\infty} \sum_{k=0}^{\infty} \frac{k'}{z} p(\mathbf{k}) \sum_{m'=0}^{k'-1} \sum_{m=0}^{k} B_m^{k'-1}(q') B_m^k(q) F(\mathbf{m}, \mathbf{k}) \tag{10}$$

Since we have assumed that the response thresholds in both layers are drawn from the same distribution, and by doing so making $F(\mathbf{m}, \mathbf{k})$ symmetric in the variables both network layers, we observe that

$$g(q, q') = g'(q', q). \tag{11}$$

Thus if we assume $q_0 = q'_0 = \rho_0$, we have $q_n = q'_n$ for all n, and we only need to concern ourselves with one equation:

$$g(q) = g(q, q) = \rho_0 + (1 - \rho_0) \sum_{k=1}^{\infty} \sum_{k'=0}^{\infty} \frac{k}{z} p(\mathbf{k}) \sum_{m=0}^{k-1} \sum_{m'=0}^{k'} B_m^{k-1}(q) B_{m'}^{k'}(q) F(\mathbf{m}, \mathbf{k}). \tag{12}$$

Moreover the fact that $F(\mathbf{m}, \mathbf{k})$ is separable into $F(\mathbf{m}, \mathbf{k}) = \Phi(m/k)\Phi(m'/k')$, we can make one final simplification:

$$\rho = \rho_0 + (1 - \rho_0) \sum_{\mathbf{k}=0}^{\infty} p(\mathbf{k}) \left(\sum_{m=0}^{k} B_m^k(q_\infty) \Phi(m/k) \right)^2 \tag{13}$$

$$g(q) = \rho_0 + (1 - \rho_0) \sum_{k=1}^{\infty} \sum_{k'=0}^{\infty} \frac{k}{z} p(\mathbf{k}) \left(\sum_{m=0}^{k-1} B_m^{k-1}(q) \Phi(m/k) \right) \left(\sum_{m'=0}^{k'} B_{m'}^{k'}(q) \Phi(m'/k') \right). \tag{14}$$

4   Simulations

5 Conclusions

References

[1] P Erdös and A Rényi. "On random graphs, I". In: *Publicationes Mathematicae (Debrecen)* 6 (1959), pp. 290–297. URL: `http://www.renyi.hu/~p_erdos/Erdos.html#1959-11` (visited on 06/17/2018).

[2] James P. Gleeson and Diarmuid J. Cahalane. "Seed size strongly affects cascades on random networks". In: *Physical Review E* 75.5 (May 3, 2007). ISSN: 1539-3755, 1550-2376. DOI: `10.1103/PhysRevE.75.056103`. URL: `https://link.aps.org/doi/10.1103/PhysRevE.75.056103` (visited on 05/30/2018).

[3] Clara Granell, Sergio Gómez, and Alex Arenas. "Dynamical Interplay between Awareness and Epidemic Spreading in Multiplex Networks". In: *Physical Review Letters* 111.12 (Sept. 17, 2013). ISSN: 0031-9007, 1079-7114. DOI: `10.1103/PhysRevLett.111.128701`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.111.128701` (visited on 06/13/2018).

[4] Matthew O. Jackson and Yves Zenou. *Games on Networks*. SSRN Scholarly Paper ID 2136179. Rochester, NY: Social Science Research Network, Jan. 1, 2014. URL: `https://papers.ssrn.com/abstract=2136179` (visited on 05/09/2018).

[5] Brian Karrer and M. E. J. Newman. "Stochastic blockmodels and community structure in networks". In: *Physical Review E* 83.1 (Jan. 21, 2011). ISSN: 1539-3755, 1550-2376. DOI: `10.1103/PhysRevE.83.016107`. arXiv: `1008.3926`. URL: `http://arxiv.org/abs/1008.3926` (visited on 05/22/2018).

[6] Kyu-Min Lee, Charles D. Brummitt, and K.-I. Goh. "Threshold cascades with response heterogeneity in multiplex networks". In: *Physical Review E* 90.6 (Dec. 29, 2014), p. 062816. DOI: `10.1103/PhysRevE.90.062816`. URL: `https://link.aps.org/doi/10.1103/PhysRevE.90.062816` (visited on 06/17/2018).

[7] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. "Random graphs with arbitrary degree distributions and their applications". In: *Physical Review E* 64.2 (July 24, 2001). ISSN: 1063-651X, 1095-3787. DOI: `10.1103/PhysRevE.64.026118`. arXiv: `cond-mat/0007235`. URL: `http://arxiv.org/abs/cond-mat/0007235` (visited on 05/18/2018).

[8] Yoav Shoham and Kevin Leyton-Brown. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge: Cambridge University Press, 2008. ISBN: 978-0-511-81165-4. DOI: `10.1017/CBO9780511811654`. URL: `http://ebooks.cambridge.org/ref/id/CBO9780511811654` (visited on 03/29/2018).

[9] Duncan J. Watts. "A simple model of global cascades on random networks". In: *Proceedings of the National Academy of Sciences* 99.9 (Apr. 30, 2002), pp. 5766–5771. ISSN: 0027-8424, 1091-6490. DOI: `10.1073/pnas.082090499`. URL: `http://www.pnas.org/content/99/9/5766` (visited on 05/29/2018).