Group Segregation in Social Networks

Dominic Aits
Imperial College
London, UK
dlmaits9@gmail.com

Alexander Carver
Imperial College
London, UK
alexander.carver15@imperial.ac.uk

Paolo Turrini
University of Warwick
Coventry, UK
p.turrini@warwick.ac.uk

ABSTRACT

We study strategic behaviour in heterogeneous network formation, where agents are grouped into types and can choose to create or sever links whilst maximising their own private interest. We show the conditions under which social networks exhibit segregated behaviour by groups, as a function of the individual benefits and the costs of linking. By introducing the idea of an individual having a degree of 'tolerance' for others not of their own type, we further show that this enriched framework is able to generate sophisticated intra-group segregation, where a group can shun one of its own members due to the connections that member has. Moreover, we find through simulations that group segregation is an endemic feature and that, as the cost of linking increases, networks converging to a stable state exhibit common characteristics with growing certainty.

KEYWORDS

Social Networks, Jackson-Wolinsky Model, Group Segregation

ACM Reference Format:
Dominic Aits, Alexander Carver, and Paolo Turrini. 2019. Group Segregation in Social Networks. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 9 pages.

1 INTRODUCTION

A significant aspect of real life social networks is that they exhibit segregation: individuals who share common characteristics routinely are more closely associated to one another relative to others who do not. This has been increasingly noticeable with the advent of online social networks, particularly in connection with undesirable phenomena such as informational bubbles, opinion polarisation and fake news [17]. Understanding the formation of group segregation in social networks is therefore a timely and important undertaking.

Over the past decades social networks have emerged in the multi-agent systems community as a computational framework for the analysis of distributed interaction. A number of frameworks have been proposed to study their dynamics, with applications such as product adoption [1, 2], opinion diffusion [5, 10, 22, 23] and community detection [23, 24]. However, despite the study of segregation dating back to Schelling [20, 21], its connection to social networks still remains largely unexplored.

The model that is possibly closest in spirit to this enterprise is the Jackson-Wolinsky (JW) model [15]. Here, a number of homogeneous agents connected in a social network can choose to further connect to, or disconnect from, other agents, as a function of the cost of linking and the benefits thereof. The key instrument for analysis is the equilibrium concept of *pairwise stability* - where two agents can form a link based on joint consent, but can unilaterally delete a link if they so choose to. The main focus is on the tension between stability and efficiency - in other words, the possibility that the networks formed from private interests do not coincide with the network structures that are optimal for society as a whole. Jackson and Wolinsky's contribution is limited to homogeneous agents in a static setting. In real life social networks, though, individuals create and destroy links in line with their preferences for the other individuals' characteristics (defined as 'types' henceforth). It is therefore natural to consider heterogeneous networks where individuals have preferences over types and to establish under what conditions they polarise themselves into clusters.

Our contribution. We generalise the JW model by adding heterogeneity among individuals' types and studying connection dynamics as strategic decisions that can generate patterns of segregation. Specifically, we introduce heterogeneity into the benefits that an individual of a given type attains from their direct connections and how this is propagated through the indirect ones (defined as *value* heterogeneity henceforth). We show under what conditions stable clusters of different types emerge, as a function of individuals' benefits and the cost of connecting.

Additionally, we introduce the notion of 'tolerance', allowing agents of a given type to have different preferences for agents of other types, inducing a further level of heterogeneity. We show that these two forms of heterogeneity combined are able to generate segregation among types (both inter-type and intra-type) during the formation of a network, where a group can shun a member of its own type due to the connections that member has. Our model therefore provides a rich micro-founded theoretical framework for how segregated groups form in social networks and reflects empirical work on group formation in real-world social networks [3, 23].

Finally, we extend our analysis to a dynamic setting, designing a simulation platform to induce social network formation in line with our theoretical framework. Simulations of generated networks, over a variety of parameter distributions, highlight that social network segregation is a common characteristic, even with a very low cost of linking. Moreover, we find that as the cost of linking increases, pairwise stable networks share common statistical properties with increasing certainty. This means that while we might not be able to predict the actual structure of a social network ex-ante, we can more accurately predict its features (at high costs of linking).

Related literature. Besides the JW model, other contributions have close connections to ours. Notably, the stream of literature on homophily, i.e., the fact that individuals with similar characteristics

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.

tend to bunch together [19]. Among these are Jackson and Rogers [13], who expand the JW model to generate small-world properties by critically capping the benefits that an individual may get from the wider network at a specific path length, and are able to generate communities in the network structure as a result. Following this same line of research, the paper that is most closely related to ours is by Gallo [9], who generates pairwise stable networks that exhibit segregation and small-world properties, with the results however being driven by cost, rather than value heterogeneity, and by information asymmetry. Gallo introduces heterogeneity into the *knowledge* of the network - some agents are fully aware of the whole structure of the network, whilst others are less knowledgeable. This is, in our view, unappealing, since segregation may not necessarily be a function of the information that an agent has about the nature of the network. Our framework therefore accounts for social networks where the cost of connecting is very low and where information about the network is highly accessible - arguably both features of modern (online) social networks.

Other related papers, but somewhat further away from the scope of ours, are [4], [6], [7] and [14], which focus on non-consensual dynamics for forming and severing connections.

Introducing agent heterogeneity in strategic network formation is not new. Galeotti et al. [8] used both value and cost heterogeneity into their models. However, there are important differences between their approach and ours. First, they characterise stable networks using Nash equilibrium as the stability concept of choice. Second, they do not conduct a general study of decay in the network. Rather, they focus primarily on networks where the path length does not matter in the magnitude of benefits. Third, they opt for (direct) cost heterogeneity over (indirect) value heterogeneity.

Other literature has also explored heterogeneity in social network formation. Johnson and Gilles [16] extend the JW model by introducing spatial cost topology. Individuals are distributed across a real number line (i.e. to represent their individual varying characteristics), and the cost of directly connecting between individuals depends on the 'length' between them. Hojman and Szeidl (2003) [11] also focus on cost heterogeneity, and specifically on core-periphery star-based structures, using a refinement of the pairwise stability equilibrium concept. McBride [18] takes an entirely new approach highliting the impact of imperfect monitoring between individuals to generate an equivalent of agent heterogeneity.

However, to our best knowledge, no paper has so far analysed the role of value heterogeneity in a pairwise stable network setting with decay, combining the equilibrium analysis of social networks with the study of segregation as a property of their stable states. Our focus on value heterogeneity is key, since segregation is driven by agents' preferences over both their direct and indirect connections, applicable in social networks where the cost of connection is negligible. Additionally, our simulations provide the first attempt to draw generalisable insights into the characteristics of complex pairwise stable social networks, based on a rich theoretical framework.

2 JACKSON-WOLINSKY NETWORKS

In this section we introduce the basic notation, definitions and results from [15], which our model and results build upon.

2.1 Social Networks

Let $\phi = \{1, ..., N\}$ be a finite set of agents, each representing a node in a graph, with the edges between the nodes modelling individual connections. We define the complete graph g^{ϕ} to be the set of all subsets of ϕ of size 2. From this, we can define any possible graph on ϕ as belonging to the set $\{g \mid g \subseteq g^{\phi}\}$. To specify a connection between any two distinct nodes $i, j \in \phi$, we denote ij to represent the undirected link between nodes i and j in the relevant subset of ϕ. Therefore, if $ij \in g$, then nodes i and j are directly connected; and if $ij \notin g$, they are not.

We can now formalise the creation and destruction of links. Let $g + ij$ yield the graph generated by taking the original graph g, and adding the link ij. Similarly, let $g - ij$ yield the graph generated by taking the original graph g, removing the link ij. In other words, $g + ij = g \cup \{ij\}$; and $g - ij = g \backslash \{ij\}$.

Let $N(g) = \{i \mid \exists j \;\; s.t. \;\; ij \in g\}$ represent the set of nodes which have at least one neighbour (i.e. direct connection), and $n(g) = |N(g)|$ to be the cardinality of this set. We can further define a path in this graph. For a graph g, and nodes i_1 and i_m, if there exists a set of distinct nodes $\{i_1, i_2, ..., i_m\} \subseteq N(g)$, and $\{i_1 i_2, i_2 i_3, ..., i_{m-1} i_m\} \subseteq g$, then there exists a path between i_1 and i_m in g.

Let $\emptyset \neq g' \subseteq g$. If, for all $i \in N(g')$ and $j \in N(g')$ where $i \neq j$, there exists a path in g' between i and j - and, for any $i \in N(g')$ and $j \in N(g)$, it is such that $ij \in g$ implies that $ij \in g'$ - then we can say that g' is a *component* of g.

2.2 Strategies and equilibria

In strategic network formation, agents choose to form or sever links between each other whilst maximising their own private interests. Hence, each agent needs to have a utility, or payoff function, ascribing the value they gain from the current state of a graph g.

We first define the *value* attributable to an *entire graph* to be $v : \{g \mid g \subseteq g^{\phi}\} \rightarrow \mathbb{R}$, and for V to be the set of all possible functions v. For example, in the true utilitarian sense, a suitable candidate for v may be $v(g) = \sum_{i \in \phi} u_i(g)$ where $u_i : \{g \mid g \subseteq g^{\phi}\} \rightarrow \mathbb{R}$.

We further define an *allocation rule*, $U : \{g \mid g \subseteq g^{\phi}\} \times V \rightarrow \mathbb{R}^{\mathbb{N}}$, that determines how the value attributing to the whole graph is allocated to the individual agents (i.e. the nodes) in the graph. Hence, the function $U_i(g, v)$ represents the utility/payoff an individual i receives from a graph g with value function v. For example, in the case above where $v(g) = \sum_i u_i(g)$, a suitable candidate for $U_i(g, v)$ is $U_i(g, v) = u_i(g)$ for all $i \in \phi$.

To analyse the evolution of the network we opt for the equilibrium notion of pairwise stability. Given a value function v and an allocation rule U, g is said to be *pairwise stable* if two conditions hold:

$$\begin{aligned} \forall ij \in g : \text{BOTH} \quad & U_i(g, v) \geq U_i(g - ij, v) \\ \text{AND} \quad & U_j(g, v) \geq U_j(g - ij, v) \end{aligned} \tag{1}$$

$$\begin{aligned} \forall ij \notin g : \text{IF} \quad & U_i(g, v) < U_i(g + ij, v) \\ \text{THEN} \quad & U_j(g, v) > U_j(g + ij, v) \end{aligned} \tag{2}$$

We say that a graph g is *defeated* by g' if $g' = g - ij$ and (1) does not hold for ij under g, or if $g' = g + ij$ and (2) does not hold for ij under g.

Intuitively, pairwise stability requires individual agents to jointly consent for a link to be formed between them, yet allows each agent to destroy any existing links they may have independently of anyone else. In particular, if a link exists in a graph between two agents, then they must both benefit individually from such a link. Likewise, if an agent is strictly better off adding a link to another, then this link can only not exist if the other agent is strictly worse off.

An interesting case emerges if both agents are indifferent about maintaining a link or not. If $ij \in g$, and agents i and j are indifferent between maintaining the link or not, then condition (1) states that the link will be kept. If $ij \notin g$, then condition (2) is not violated either. Therefore, if two agents are indifferent between a link or not, then that link may, or may not, exist. For ease of simulation in Section 5 and without loss of generality, we stipulate the following additional assumption: *if two distinct agents $i, j \in \phi$ are indifferent between $g_{ij} = 1$ and $g_{ij} = 0$, then form the link such that $g_{ij} = 1$.*

So now links are formed when both agents are weakly better off with the link than without it. Equivalently, we can re-state the definition of pairwise stability, weakening condition (2) as follows:

$$\begin{aligned} \forall ij \notin g : \text{IF} \qquad & U_i(g, v) \leq U_i(g + ij, v) \\ \text{THEN} \qquad & U_j(g, v) > U_j(g + ij, v) \end{aligned} \tag{3}$$

2.3 A Homogeneous Connections Model

We can now move to define the complete Jackson-Wolinsky model, by first introducing the *general* utility function for an agent i, which we denote $U_i(g)$, representing the utility that i gains from graph g and letting $u_{ij} \in \mathbb{R}$ be the utility that i receives when directly connecting to an agent j. On top of this, let $\delta \in (0, 1)$ be the decay rate at which the utility from an indirect connection between agents i and j is discounted by, raised by the geodesic distance (i.e. shortest path length) between these two agents, denoted by $dist(i, j)$. Finally, we assume a cost of direct communication or connection between any two agents i and j, denoted by c_{ij} and such that $c_{ij} \in \mathbb{R}_+$. Therefore, the general utility function can be written as:

$$U_i(g) = \sum_{k \in \phi} \delta^{dist(i,k)} u_{ik} - \sum_{k:\ ik \in g} c_{ik} \tag{4}$$

Notice that equation (4) embodies both *value* and *cost* heterogeneity in its current formulation. u_{ik} can differ among different pairings of i and k, allowing for heterogeneity among the benefits that individuals can gain from connections, either direct or indirect. Similarly, c_{ik} can differ among pairings of i and k as well, representing that there may be different costs of connection for different individuals. Further, note that every individual receives a benefit u_{ii}, without any discount, irrespective of what connections they have. If there is no path between i and j, then $dist(i, j) \to \infty$, with the result that no benefit is attained to either node from a lack of such a path.

Jackson and Wolinsky place a series of constraints on equation (4) to generate a homogeneous agent model. In particular, they assume symmetry, with $c_{ij} = c$ for all ij (hence cost homogeneity), and $u_{ij} = 1$ for all ij (hence value homogeneity).

They then proceed to prove the following proposition for a characterisation of pairwise stable networks, which we state for reference:

Proposition 2.1. *(Jackson and Wolinksy (1996) [15]). Assume a symmetric connections model with $U_i(g) = u_i(g)$. Then the following facts are true:*

(i) A pairwise stable network has at most one component.

(ii) For $c < \delta - \delta^2$, the unique pairwise stable network is the complete graph, g^{ϕ}.

(iii) For $\delta - \delta^2 < c < \delta$, a star encompassing all players is pairwise stable, but not necessarily the unique pairwise stable graph.

(iv) For $\delta < c$, any pairwise stable network which is non-empty is such that each player has at least two links and thus is inefficient.

Whilst the homogeneous model is powerful in the sense that it is capable of generating a relatively small subset of pairwise stable networks over large cost ranges, all agents belong to the same type. Therefore it is incapable of explaining how segregation occurs between different groups. Hence, it is necessary to extend it with heterogeneous agents, which we proceed to do next.

3 HETEROGENEOUS DYNAMICS

The study of segregation in social network formation requires agents to have a ‘type’ associated with themselves. In our model types are simply encoded as elements $t_i \in [0, 1]$, one for each agent i, with T being the set of all available types. We then redefine the payoff function that allocates a utility value to agent i with type t_i from a direct connection to another agent j with type t_j under graph g as $u_i(t_i, t_j, g) \in \mathbb{R}$, and specify that the allocation rule generates a utility for the agent i under the standard rule:

$$U_i(g) = \sum_{k \in \phi} \delta^{dist(i,k)} u_i(t_i, t_k, g) - \sum_{k:\ ik \in g} c_{ik} \tag{5}$$

Our formulation allows for a number of natural assumptions to be placed upon the nature of $u_i(t_i, t_j, g)$. In particular, we will assume that every agent would prefer to connect with individuals who have a type closer to their own. Thus:

$$u_i(t_i, t_j, g) < u_i(t_i, t_k, g) \text{ where } t_j < t_k \leq t_i \tag{6}$$

$$u_i(t_i, t_j, g) < u_i(t_i, t_k, g) \text{ where } t_j > t_k \geq t_i \tag{7}$$

$$u_i(t_i, t_j, g) = u_i(t_i, t_k, g) \text{ where } t_j = t_k \tag{8}$$

We further impose cost homogeneity for all agents as per [15]. To showcase our model, we now proceed to state and prove some propositions, focussing once more on the characterisation of pairwise stable networks. To more neatly illustrate the framework's flexibility in generating varying pairwise stable architectures under different parameter constraints, we choose to restrict ourselves to networks with only two types T_1 and T_2 which, notice, can be naturally lifted to richer structures. In what follows, we describe patterns of segregations in case the discounted utility of connecting across types is higher than its cost (Proposition 3.1), or not

(Proposition 3.2). Then we describe the exact structure of segregated equilibria when connecting across types induces disutility, with only two types (Proposition 3.3) or more (Theorem 3.6).

Proposition 3.1. *Assume a heterogeneous connections model with types $T = \{T_1, T_2\}$ drawn from the $[0, 1]$ interval. Further assume that both $\delta u(T_1, T_2, g) > c$ and $\delta u(T_2, T_1, g) > c$. Finally, w.l.o.g., assume that $u(T_1, T_2, g) \geq u(T_2, T_1, g)$. Then:*

(i) For $c < (\delta - \delta^2)u(T_2, T_1, g) \leq (\delta - \delta^2)u(T_1, T_2, g)$, the complete network is the unique pairwise stable network.

(ii) If $(\delta - \delta^2)u(T_1, T_2, g) < c < (\delta - \delta^2)u(T_1, T_1, g)$ and $(\delta - \delta^2)u(T_2, T_1, g) < c < (\delta - \delta^2)u(T_2, T_2, g)$, then any pairwise stable equilibrium will have complete connections between members of the same type. If $c < (\delta - \delta^3)u(T_1, T_2, g)$ and $c < (\delta - \delta^3)u(T_2, T_1, g)$, each member of a type has a unique, and single, connection to a member of the opposite type, if one is available. If $c > (\delta - \delta^3)u(T_1, T_2, g)$ and $c > (\delta - \delta^3)u(T_2, T_1, g)$, then there is only one connection between the two types.

(iii) If $(\delta - \delta^2)u(T_1, T_1, g) < c < \delta u(T_1, T_2, g)$ and $(\delta - \delta^2)u(T_2, T_2, g) < c < \delta u(T_2, T_1, g)$, a star, with the centre of any type, is a pairwise stable network, but it may not be unique.

Proof. (i) By $c < (\delta - \delta^2)u(T_2, T_1, g) \leq (\delta - \delta^2)u(T_1, T_2, g)$, we have that $c < (\delta - \delta^2)u(T_1, T_1, g)$ and $c < (\delta - \delta^2)u(T_2, T_2, g)$. Consider a graph g in which an agent i of type T_1 is not directly connected to an agent j of type T_2. For agent i, if we form the link ij, then i gets a benefit of at worst $(\delta - \delta^2)u(T_1, T_2, g) - c$ (in giving up an indirect connection to j). Therefore $u^{ij} = u(T_1, T_2, g + ij) - u(T_1, T_2, g)$ is strictly positive. Similarly the same applies to j with $u^{ji} > 0$. The same analysis also follows for i and j if they are of the same type. Therefore there cannot be any pairwise stable graph g which is not completely connected.

(ii) By $c < (\delta - \delta^2)u(T_1, T_1, g)$, all agents of type T_1 are directly connected to one another. The same applies to agents of type T_2 since $c < (\delta - \delta^2)u(T_2, T_2, g)$. Assume that there exists a pairwise stable graph g where two agents of the same type, say k and l of type T_1 w.l.o.g, both have a direct connection to an agent of type T_2, say m. We know that k and l are directly connected. If k chooses to delete their link with m, then they get a marginal payoff of $c - (\delta - \delta^2)u(T_1, T_2, g) > 0$, so there will be no link between k and m. If there are more than two agents of type T_1 connected to the same agent of type T_2, the same argument applies. Equally, if k and l were of type T_2, and m of type T_1, the same argument applies. Thus, there can be at most one direct connection between agents of different types. Moreover, no agent of a single type will have more than one connection to the component consisting of agents of the other type, since generating the second connection will yield a marginal utility of $(\delta - \delta^2)u(T_1, T_2, g) - c < 0$ or $(\delta - \delta^2)u(T_2, T_1, g) - c < 0$ as the direct connection will forego the benefits of an indirect connection valued at δ^2. Finally, we show that a graph g consisting of two separate complete networks of the individual types, non-overlapping, is not pairwise stable. If $c < (\delta - \delta^3)u(T_1, T_2, g)$ and $c < (\delta - \delta^3)u(T_2, T_1, g)$, then, w.l.o.g, an agent i, of type T_1, can directly connect to an agent of the other component of type T_2 and gain a minimum of $(\delta - \delta^3)u(T_1, T_2, g) - c > 0$. The same applies for agents of the other type. Thus, the unique pairwise stable network is with all agents of the same type directly connected to one another, and where every agent of this type, if a connection is available, will have a unique direct connection to an agent of the other type. If $c > (\delta - \delta^3)u(T_1, T_2, g)$ and $c > (\delta - \delta^3)u(T_2, T_1, g)$, then since $c < \delta u(T_1, T_2, g)$ and $c < \delta u(T_2, T_1, g)$, there will be only one connection between the two types.

(iii) Consider a graph g that is a star, i.e. one agent is the centre node, and all other agents have a direct connection to the central agent. W.l.o.g say that the central agent is of type T_1. This agent has no incentive to delete links to either agents of the other type (since $c - \delta u(T_1, T_2, g) < 0$), or to agents of the same type (since $c - \delta u(T_1, T_1, g) < c - \delta u(T_1, T_2, g) < 0$). Consider a peripheral agent, w.l.o.g, of type T_1. This agent has no incentive to connect to another periphery agent of either type, since $(\delta - \delta^2)u(T_1, T_2, g) - c < 0$ and $(\delta - \delta^2)u(T_1, T_1, g) - c < 0$. Obviously, a periphery agent will not want to delete the link with the centre agent irrespective of its type since $c < \delta u(T_1, T_2, g) < \delta u(T_1, T_1, g)$. Therefore, the star is a pairwise stable network. □

Propositions (3.1)(i) and (3.1)(ii) show that the introduction of heterogeneity still allows for unique pairwise equilibria. In particular, we are able to showcase that complete integration of types is still possible as unique equilibrium, albeit under restrictive cost ranges.

Proposition (3.1)(iii) allows for multiple pairwise stable equilibria. The star, for example, with either type at its centre, is a pairwise stable equilibrium. What this highlights is that in networks with different types, and where types are in communication with one another (i.e. there is a degree of integration), the resulting network structure may exhibit a degree of centrality. These networks may have particular agents who act as 'gatekeepers' to other individuals of the same type, and all others have to communicate through these gatekeepers to gain access to others. This feature is indicative of a small-worlds property common in many social networks – see for instance [12].

Proposition (3.1) is also interesting in that no agent receives any disutility from connections, either direct or indirect, with agents of the other type. Regardless, we still do not get complete integration between communities. This is because of the presence of externalities, which is also typical of the homogeneous connections model. Agents attempt to economise on their direct connections if the cost of connecting is too high. Instead they rely on the positive benefits (i.e. externalities) that their indirect connections provide.

For completeness, we also consider the case where a direct connection to a member of the other type yields a cost instead of a net gain. It is important to notice that the proposition below does not rely on an agent receiving disutility from an indirect connection to an agent of the other type.

Proposition 3.2. *Assume a heterogeneous connections model with types $T = \{T_1, T_2\}$ drawn from the $[0, 1]$ interval. Further assume that $c > \delta u(T_1, T_2, g)$ and $c > \delta u(T_2, T_1, g)$. Then, for any agent of a given type, if they have a direct connection to a second agent of another type, then this second agent must have at least two links in any pairwise stable network.*

Proof. This is a complementary proposition to Proposition (2.1)(iv). Suppose a non-empty network g is pairwise stable whereby an agent i, w.l.o.g, of type T_1 has a direct connection to an agent j of type T_2. Suppose further that j's only direct connection is with i. Then i can strictly benefit by deleting its link with j and receive $c - \delta u(T_1, T_2, g) > 0$. Hence g is not pairwise stable. □

The presence of agents of different types in the same component of a pairwise stable networks is entirely possible. For example, the net cost of a direct connection between agents of different types may be outweighed if an agent of a specific type brings enough indirect connections to be worth connecting to. However, the driver of such network formation lies primarily in the cost of connection, and impacts all types equally.

The true value of heterogeneous models is to assess situations where an agent actually gains disutility from a connection, either direct or indirect, to an agent of a differing type. The important question is then whether all non-empty pairwise stable networks are truly segregated, in the sense that agents of a given type reside in their own distinct components. The following proposition characterises the nature of pairwise stable networks under this framework.

Proposition 3.3. *Assume a heterogeneous connections model with types $T = \{T_1, T_2\}$ drawn from the $[0, 1]$ interval. Assume further that both $0 > u(T_1, T_2, g)$ and $0 > u(T_2, T_1, g)$. Finally, w.l.o.g, assume that $u(T_1, T_1, g) \geq u(T_2, T_2, g)$. Then:*

(i) Any non-empty acyclic pairwise stable network consists of at most two separate components. There is no path in either component which contains agents of both types. Each component is homogeneous in terms of types.

(ii) If $c < (\delta - \delta^2)u(T_2, T_2, g) \leq (\delta - \delta^2)u(T_1, T_1, g)$, the unique acyclic pairwise stable network consists of two separate components, each only consisting of agents of the same type, completely connected.

Proof. (i) Suppose an acyclic pairwise stable network g has at least one non-trivial component. Further suppose that one of these components contains agents of types T_1 and T_2. By definition, an agent i of type T_1 must have a direct connection to an agent j of type T_2. For the link ij to occur in a pairwise stable network, then i must be gaining an indirect benefit from its connection to j since the direct connection to j yields a disutility of $\delta u_i(T_1, T_2, g) - c < 0$ (and similarly for j as well). This indirect utility is only attainable from a component called B_1 accessed through the shortest path via j (otherwise i would gain the indirect benefit from another path). Denote the utility of the component B_1 as U_{B_1} via j. Then the utility for i from link ij is $\delta^2 U_{B_1} + \delta u_i(T_1, T_2, g) - c > 0$ since the component B_1 is at a path length of 2 away from i.

The component B_1 must contain at least one agent of type T_1, say k_1, that has at least one direct connection to an agent of type T_2. By Proposition (3.2), k_1 must have at least two links. An agent of type T_2 (in this case j) will only want to directly connect to k_1 if there is a component accessible only through k_1 which contains agents of type T_2. Call this component B_2, and let an agent of type T_2 in B_2 be m_2 (to which k_1 is connected to). Moreover, $B_2 \subset B_1$ since g is acyclic. Similarly k_1 will only want to maintain a connection to component B_2 if it contains agents of type T_1 that are directly connected to m_2. These agents of type T_1 themselves will only want to directly connect to m_2 if there is a component $B_3 \subset B_2$, accessible only through m_2, which contains agents of type T_1.

Thus the problem repeats itself over and over again, without termination. As the graph has a finite number of agents, at some point an agent of, say w.l.o.g, type T_1 will directly connect to a component (or singleton) only consisting of type T_2. Thus any links between these agents will be severed, and links between types T_1 and T_2 will be severed recursively. Thus g cannot be pairwise stable where there is a path containing agents of different types.

(ii) By Proposition (3.3)(i) we know that agents of the same type will all reside in the same component with no other agents of the other type for any acyclic graph. If $c < (\delta - \delta^2)u(T_2, T_2, g) \leq (\delta - \delta^2)u(T_1, T_1, g)$, then two applications of Proposition (2.1)(ii) to the two components individually will generate completely connected components. □

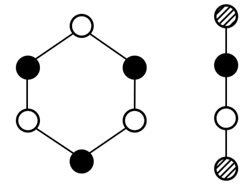

Figure 1: On the left, a cycle consisting of two types, which is pairwise stable under conditions provided by Example 3.4. On the right, a line network consisting of three types, which is pairwise stable as per the conditions in Example 3.7.

Proposition (3.3) has remarkable consequences. In particular, it predicts that if there is any disutility between two different types in a heterogeneous connections model (with only two types), then complete segregation between the two types, for any cost range, will be a feature in acyclic non-trivial pairwise stable networks.

Note however that in cyclic graphs, agents of either types can still have a direct connection, as the following counter example shows.

Example 3.4. Consider 6 agents, 3 of type T_1 and 3 of type T_2, as depicted in Figure 1 (left side). Moreover, suppose that each agent has two connections to agents of the other type only (hence a cycle of alternating types). Consider an agent i of type T_1. i will not want to delete one of its two links if $c - \delta u_i(T_1, T_2, g) - \delta^2 u_i(T_1, T_1, g) + \delta^4 u_i(T_1, T_1, g) + \delta^5 u_i(T_1, T_2, g) < 0$. i will further not choose to link with another agent of type T_1 if $c > \delta u_i(T_1, T_1, g) - \delta^2 u_i(T_1, T_1, g) + \delta^2 u_i(T_1, T_2, g) - \delta^3 u_i(T_1, T_2, g)$. These provide bounds on c. Some algebraic steps the yields the inequality $u_i(T_1, T_1, g)(2\delta - 1 - \delta^3) > u_i(T_1, T_2, g)(\delta(1-\delta) - (1-\delta^4))$. The RHS of this inequality is positive as $u_i(T_1, T_2, g) < 0$ and $1-\delta^4 = (1-\delta^2)(1+\delta^2) = (1-\delta)(1+\delta)(1+\delta^2)$. Thus $(1 - \delta)(1 + \delta)(1 + \delta^2) > (1 - \delta) > \delta(1 - \delta)$. Suppose that $u_i(T_1, T_2, g)$ is extremely close to 0. Consider $2\delta - 1 - \delta^3$. In the positive domain for δ this expression has real roots at $\delta = 1$ and $\delta = \frac{\sqrt{5}}{2} - \frac{1}{2}$ with the expression itself being positive between these two real roots. Thus the cost range for c is valid, and the cycle is pairwise stable (the same analysis can be conducted for agents of the other type as the problem is symmetric in approach).

The reason for why the proof for Proposition (3.3) does not carry through to cyclic graphs is due to the refining of component sizes. For two agents that provide disutility to one another to directly connect, there must be components linked to these agents that provide both of them with utility. Our proof above shows that, with only two types in an acyclic network, there is no stable component that is not a singleton that allows this to occur. We reduce the size of these components since the graph is acyclic as the proof iterates until there are no more agents left in the network. If the graph is cyclic, this argument does not hold as the counter example above shows.

A key question therefore arises: whether it is possible to connect types, which generate disutility to one another, in a heterogeneous acyclic network with more than two types. We first fix a useful definition.

Definition 3.5. (*Gross disutility*) An agent i with type T_1 is said to provide gross disutility to agent j with type T_2 if $u_j(T_2, T_1, g) < 0$.

We now state a general theorem which enables the existence of acyclic pairwise stable networks to occur where agents of different types who provide gross disutility to one another are directly connected.

Theorem 3.6. *Let the graph g be acyclic. Assume a heterogeneous connections model with types given by $T = \{T_1, ..., T_m\}$, drawn from the $[0, 1]$ interval. Consider two types, where either or both provides gross disutility to one another. Suppose further that there is a direct connection between agents of these two types, say, w.l.o.g, between agents i of type T_1 and j of type T_2. Let A be the component accessible via i, and B be the component accessible via j. Then in any pairwise stable network, if i faces gross disutility from its connection to j, then B must contain at least one agent of a different type, w.l.o.g say T_3, that provides utility to i. Similarly, the same applies for j with component A.*

Proof. This is a repeated application of Proposition (3.3). Consider an agent i of type T_1 directly connected to agent j of type T_2. Let A be the component accessible via i, and B be the component accessible via j. Suppose further that i gets gross disutility from its connection to j. If B contains only agents of type T_2, then i will sever its connection to j. If B only contains agents of type T_1, then j will want to severe its connection to B. If B contains agents of types T_1 and T_2, then by Proposition (3.3), g cannot be pairwise stable. Therefore, B must contain at least one agent of a separate type T_3 that provides utility to i. The same analysis can be applied to A. □

Theorem 3.6 highlights the importance of indirect connections in fostering links between types that yield gross disutility to one another. An agent i may choose to tolerate a connection to a second agent j that provides gross disutility to i if j has valuable connections with high enough benefits for i. In other words, a greater variety of types can mean a pairwise stable graph is better connected. The following example showcases the mechanism behind Theorem 3.6.

Example 3.7. Consider a graph g, with three types T_1, T_2, and T_3, as depicted in Figure 1 (right side). There are 4 agents: a and b of type T_3, c of type T_1 and d of type T_2. c and d provide gross disutility to one another. Consider a chain where a is directly connected to c, c to d, and d to b. Assume for simplicity that $u_a(T_3, T_1, g) = u_b(T_3, T_1, g) = u_a(T_3, T_2, g) = u_b(T_3, T_2, g)$ and assume symmetry via $u_a(T_3, T_3, g) = u_b(T_3, T_3, g)$. Further, assume $u_c(T_1, T_1, g) = u_d(T_2, T_2, g)$ and $u_c(T_1, T_2, g) = u_d(T_2, T_1, g)$. As a final step, assume $u_c(T_1, T_3, g) = u_d(T_2, T_3, g)$.

Let us first consider agent a (analysis is the same for b). a will not delete its link if:

$$c - \delta u_a(T_3, T_1, g) - \delta^2 u_a(T_3, T_2, g) - \delta^3 u_a(T_3, T_3, g) < 0 \quad (9)$$

a will not add a link to d if:

$$(\delta - \delta^2) u_a(T_3, T_2, g) + (\delta^2 - \delta^3) u_a(T_3, T_3, g) - c < 0 \quad (10)$$

a will not add a link to b if:

$$\delta u_a(T_3, T_3, g) - \delta^3 u_a(T_3, T_3, g) - c < 0 \quad (11)$$

A close inspection reveals that equation (11) dominates equation (10), since $u_a(T_3, T_3, g) > u_a(T_3, T_2, g)$. Equations (9) and (11) are consistent if $\frac{u_a(T_3, T_1, g)}{u_a(T_3, T_3, g)} > \frac{1-2\delta^2}{1+\delta}$. Now consider c (the analysis is the same for d). c will not delete its link with a if:

$$c - \delta u_c(T_1, T_3, g) < 0 \quad (12)$$

c will not delete its link with d if:

$$c - \delta u_c(T_1, T_2, g) - \delta^2 u_c(T_1, T_3, g) < 0 \quad (13)$$

c will not connect with b if:

$$\delta u_c(T_1, T_3, g) - \delta^2 u_c(T_1, T_3, g) - c < 0 \quad (14)$$

Observe now that equation (13) dominates equation (12), since $\delta < 1$ and $u_c(T_1, T_2, g) < 0$. Finally, notice that equations (13) and (14) are consistent if $2\delta - 1 > -\frac{u_c(T_1, T_2, g)}{u_c(T_1, T_3, g)} > 0$. This completes the stability analysis by symmetry. For example, such a network is pairwise stable if the gross disutility between c and d is very small, and the utility that a and b get from c (and therefore d) is similar to the direct utility they would get by directly connecting themselves.

4 ADDING TOLERANCE

There are natural limitations to the heterogeneous connections model. In real world examples we see networks bifurcated with two distinct parties still communicating with one another. For example, two political parties may be opposed from an official policy point of view, yet some members of either party may occasionally speak to one another. Naturally these networks do not always feature cycles. We therefore need a richer framework than the one we have currently designed so far.

Our solution is to introduce heterogeneity between members of the same type. Consider for example two agents of a given type: one such member may choose to shun other types due to the gross disutility associated with a direct connection. However, the other member may be tolerant of some other types, and so might be willing to connect to some other types. Therefore, we introduce the idea that members of the same types may have varying degrees of tolerance for members of other types.

The way this can be achieved is by adjusting the utility function that dictates the utility agent i gets from its connection to agent j. In the previous section we defined this as $u_i(T_1, T_2, g)$ where T_1 is the type of i and T_2 is the type of j. If we want to introduce heterogeneity within types, we need an additional parameter for this functional form. We define the *degree of tolerance* as $\sigma \in [0, 1]$. Thus, the utility function is now defined as $u_i(T_1, T_2, \sigma_i, g)$.

We further posit the following relational features for the utility functions with respect to the degree of tolerance, that says that an agent's utility from a connection to another is weakly increasing as the agent's degree of tolerance increases:

$$u_i(t_i, t_j, \sigma_i + \epsilon, g) \geq u_i(t_i, t_j, \sigma_i, g) \;\; \forall \epsilon > 0 \tag{15}$$

We further provide additional definitions regarding the impact of the tolerance in this more general framework.

Definition 4.1. (*Minimum absolute tolerance (MAT)*). Let the set of degrees of tolerance, for which an agent i of type T_i will have a non-negative utility from a connection to any other agent, be $\beta_i = \{ \sigma_i \mid u_i(T_i, T_j, \sigma_i, g) \geq 0, \forall T_j \in T \}$. The *minimum absolute tolerance* ${\sigma_i}^{MAT}$ for agent i is the lowest degree of tolerance in the set β_i for which this is true. In other words, ${\sigma_i}^{MAT} = inf(\beta_i)$.

Definition 4.2. (*Minimum absolute intolerance (MAI)*). Let the set of degrees of tolerance, for which an agent i of type T_i will have a negative utility from a connection to any other agent other than to agents of its own type T_i be:

$$\alpha_i = \left\{ \sigma_i \;\middle|\; \begin{Bmatrix} u_i(T_i, T_i, \sigma_i, g) \geq 0 \\ u_i(T_i, T_j, \sigma_i, g) < 0 \; if \; T_i \neq T_j \end{Bmatrix} \right\}$$

The *minimum absolute intolerance* ${\sigma_i}^{MAI}$ for agent i is the highest degree of tolerance in the set α_i for which this is true. In other words, ${\sigma_i}^{MAI} = sup(\alpha_i)$.

The two definitions above allow us to showcase a pairwise stable network with two types, where agents of the same type shun one of their own members. This is achieved by using differing levels of tolerance between types and also within a type (with N nodes each). Suppose that all agents of type T_2 have a high degree of tolerance, say the minimum absolute tolerance level, MAT. Further suppose that $(N - 1)$ agents of type T_1 have a degree of tolerance equal to the minimum absolute intolerance level, MAI, with the remaining agent of type T_1 having a degree of tolerance of MAT. Assume, w.l.o.g., that the cost of connection is zero. Finally assume that $|N\delta u_i(T_1, T_2, {\sigma_i}^{MAI}, g)| > |u_i(T_1, T_1, {\sigma_i}^{MAI}, g)|$. Recall that the definition of MAI means that agents of a specific type gets disutility from its connections with an agent of a separate type. Then a pairwise stable network is one where all nodes of type T_2 are completely connected, and are also completely connected with the one node of type T_1. The remaining nodes of type T_1 are in their own separate component, completely connected. They do not want to connect to the lone agent of type T_1 since the indirect connections to agents of type T_2 generates enough disutility such that they would rather not connect.

5 EXPERIMENTAL RESULTS

The introduction of tolerance into the heterogeneous connections model has added further complexity into the framework, when results are highly dependent on an increasingly large parameter space that considers not only the types of agents, but also their tolerance level for other types. Whilst further theoretical analysis is certainly desirable, it is useful to explore the nature of pairwise stable networks by computer-aided simulation.

We set up a platform to simulate a dynamic network formation process, at varying costs of linking to generate statistical insights. This requires three components: the starting state of the network, the terminating conditions, and the process by which links are formed or severed. We begin every network formation process with the empty network (i.e. a network full of singleton nodes) since all link formation must be based as per the rules according to our framework.

As for link creation or destruction between nodes, we proceed as follows. We test every *possible* edge that the network could have (i.e. $\frac{N(N-1)}{2}$ checks for N nodes) according to equations 1 and 3. If a link has not been formed for the edge under consideration, we assess whether both agents would rather form the link; if the link has been formed prior, we assess whether either agent is strictly better off without it. Based on this, we appropriately add or delete links between nodes. The order of potential edges to assess whether a link should exist or not is randomly selected.

We consider two terminating conditions. The first case is that no links are created or severed when all potential edges are assessed - i.e. the network is pairwise stable. The other case is that the network has reached a state that it has previously visited in the formation process, in other words, a cycle. Whilst the existence of a cycle is not sufficient to say that a network will never reach a pairwise stable state, a priori we are unable to know this, so we terminate the algorithm upon detecting a cycle to avoid cases where a pairwise stable state might not exist at all.

We consider a three-type model, with 5 nodes each of types 0, 0.5 and 1. The decay rate is set at 0.9. We initialise the tolerance values as per the uniform distribution on the unit interval since we want to assess how networks form given various tolerance values. We consider costs of linking from 0.1 to 0.9 at 0.1 increments, with 10,000 simulations per increment. The utility function takes the functional form below, encapsulating symmetry in preferences and allowing for disutility if two agents are of sufficiently different types:

$$u_i(T_i, T_j, \sigma_i, g) = exp\left(-\frac{1}{2} \left(\frac{T_i - T_j}{\sigma_i} \right)^2 \right) - 0.2 \tag{16}$$

Over the simulated networks, we calculate the mean and variance of the number of cliques and degree centrality. These measures capture how directly connected the network is. To capture the concept of indirect connections, we design our own segregation measure where T_i refers to a type drawn from the $[0, 1]$ interval and T_i' to be the set of agents with type T_i:

$$S_{T_i,T_j} = \begin{cases} \frac{1}{|T_i'|} \sum\limits_{k \in T_i'} \sum\limits_{l \in T_j' \setminus \{k\}} \frac{1}{dist(k,l)} \cdot \frac{1}{|T_i'|-1} & \text{if } T_i = T_j \\ \frac{1}{|T_i'|} \sum\limits_{k \in T_i'} \sum\limits_{l \in T_j'} \frac{1}{dist(k,l)} \cdot \frac{1}{|T_j'|} & \text{if } T_i \neq T_j \end{cases} \tag{17}$$

Our segregation measure is an inverse measure, bounded below by 0 (for the empty network) and above at 1 (for the complete

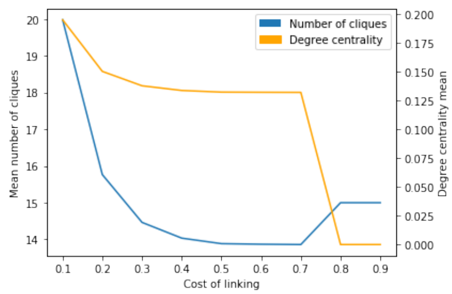

Figure 2

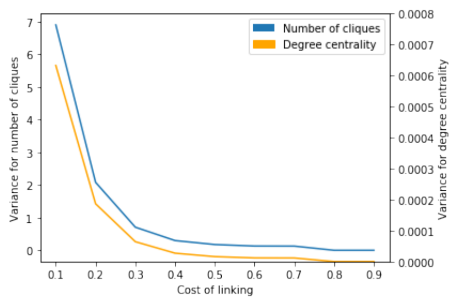

Figure 3

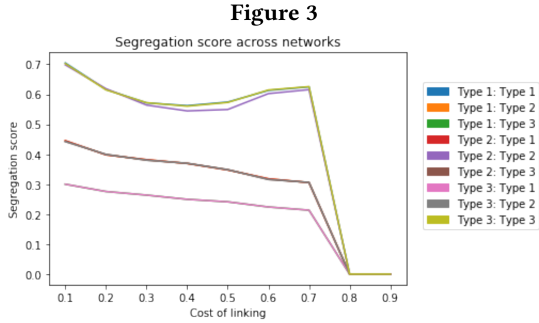

Figure 4

network), and is applicable for any pairwise stable network architecture. Moreover, it captures the notion of *intra-group* segregation (for S_{T_i,T_i}) as well as *inter-group* segregation (for S_{T_i,T_j} and $i \neq j$).

The computational results yield three main findings. With regard to the number of cliques and degree centrality, unsurprisingly, we find a negative correlation between the average number of cliques and the cost of linking - as the cost of linking gets more expensive, the network is less directly connected (above 0.7 the empty network is the only pairwise stable network, hence the number of cliques is 15 and degree centrality is 0). This can be observed in Figure 2. The more surprising finding is that the variance in the number of cliques drops markedly as the cost of linking increases, as featured in Figure 3. At low costs of linking in particular, if there is a marginal increase in cost, then the variance in the number of cliques drops by far more than at higher costs of linking. This feature is consistent with degree centrality as well.

This suggests that at low costs of linking (exempting complete networks and cost ranges where unique pairwise stable networks can be derived), there are more permissible pairwise stable networks that could have a variety of architectures. What it also suggests is that if we observe a wide range of network structures across many separate samples for a given scenario, then the cost of linking is likely to be low. As the cost of linking increases, we can better identify the statistical features of any pairwise stable network. Based on these results we cannot say with certainty that there are fewer pairwise stable networks as the cost of linking increases - but we can say that such pairwise stable networks share increasingly more similar statistical features as the cost of linking increases. Even though we might not be able to predict the exact structure of a pairwise stable network, we can say, with a high degree of certainty, what features it can exhibit, in situations where the cost of linking is known to be relatively high.

We finally show all 9 possible segregation scores in Figure 4. As the cost of linking increases, we see a reduction in all segregation scores, implying that inter-group and intra-group segregation is increasing (recall that our score metric is an inverse measure). However, beyond a cost of linking of 0.4 - 0.5, the intra-group segregation score actually increases to the detriment of the inter-group segregation score. This suggests agents of a particular type can substitute their inter-type connections in favour of intra-type connections at high costs of linking. A variety of types might not lead to more integration in a social network if the cost of linking is high.

6 CONCLUSION

We have developed a model of strategic behaviour in social network formation, showing analytical conditions under which clustering emerges among individuals belonging to different types. It is achieved via two kinds of heterogeneity - both across types of agents, but also within the same type of agent through what we define as 'tolerance' of others. Our theoretical model is able to generate both intra-group and inter-group segregation, and accounts for the impact of both direct and indirect connections.

We have further extended the model to a dynamic setting, with a surprising finding - that as the cost of linking increases, pairwise stable networks increasingly share the same statistical characteristics with increasing certainty. This suggests that while we may not be able to predict ex-ante the end state of a social network, depending on the cost of linking, we may be more able to predict its underlying features and characteristics more precisely.

Our paper has multiple avenues for expansion. First large network simulations with many agents of multiple types would be a worthy exercise, to see if the findings found here adjust with the network size. Second, the model presented here is one of perfect information. In reality, agents have imperfect information of parts of the network which they are connected to, which motivates analysis of an imperfect information model. Finally, this model presents a powerful way to explore how diffusion of ideas and habits propagate through social networks. The model can be extended to a two-stage game: in the first stage agents connect to one another as in this paper; in the second stage, their types can be adjusted depending on who they are connected to, and thus influenced by.

REFERENCES

[1] Krzysztof R. Apt and Evangelos Markakis. 2014. Social Networks with Competing Products. *Fundam. Inform.* 129, 3 (2014), 225–250.

[2] Krzysztof R. Apt, Evangelos Markakis, and Sunil Simon. 2016. Paradoxes in social networks with multiple products. *Synthese* 193, 3 (2016), 663–687.

[3] Eytan Bakshy, Solomon Messing, and Lada A. Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348 (2015), 1130–1132.

[4] Venkatesh Bala and Sanjeev Goyal. 2000. A Noncooperative Model of Network Formation. *Econometrica* 68, 5 (2000), 1181–1229.

[5] Robert Bredereck and Edith Elkind. 2017. Manipulating Opinion Diffusion in Social Networks. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI'17)*. AAAI Press.

[6] L. Elisa Celis and Aida Sadat Mousavifar. 2016. A Model for Social Network Formation: Efficiency, Stability and Dynamics. *Unpublished* (2016).

[7] Sergio Currarini, Matthew O. Jackson, and Paolo Pin. 2009. An Economic Model of Friendship: Homophily and Segregation. *Econometrica* 77, 4 (2009), 1003–1045.

[8] Andrea Galeotti, Sanjeev Goyal, and Jurjen Kamphorst. 2006. Network formation with heterogeneous players. *Games and Economic Behaviour* 54, 2 (2006), 353–372.

[9] Edoardo Gallo. 2009. Small world networks with segregation patterns and brokers. *Unpublished* (2009).

[10] Umberto Grandi, Emiliano Lorini, and Laurent Perrussel. 2015. Propositional Opinion Diffusion. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems (AAMAS '15)*. 989–997.

[11] Daniel Hojman and Adam Szeidl. 2003. Core and Periphery in Endogenous Networks. *Mimeo, Harvard University* (2003).

[12] Matthew O. Jackson. 2010. *Social and Economic Networks*. Princeton University Press.

[13] Matthew O. Jackson and Brian W. Rogers. 2005. The economics of small worlds. *Journal of the European Economic Association* 3, 2-3 (2005), 617–627.

[14] Matthew O. Jackson and Brian W. Rogers. 2007. Meeting Strangers and Friend of Friends: How Random are Social Networks. *The American Economic Review* 97, 3 (2007), 890–915.

[15] Matthew O. Jackson and Asher Wolinsky. 1996. A Strategic Model of Social and Economic Networks. *Journal of Economic Theory* 71, 0108 (1996), 44–74.

[16] Cathleen Johnson and Robert P. Gilles. 2000. Spatial Social Networks. *Review of Economic Design* 5, 3 (2000), 273–299.

[17] Jae Kook Lee, Jihyang Choi, Cheonsoo Kim, and Yonghwan Kim. 2014. Social Media, Network Heterogeneity, and Opinion Polarization. *Journal of Communication* 64, 4 (2014), 702–722.

[18] Michael McBride. 2006. Imperfect monitoring in communication networks. *Journal of Economic Theory* 126 (2006), 97–119.

[19] Miller McPherson, Lynn Smith-Lovin, and James M Cook. 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* 27 (2001), 415–444.

[20] Thomas C. Schelling. 1969. Models of Segregation. *The American Economic Review* 59, 2 (1969), 488–493.

[21] Thomas C. Schelling. 1971. Dynamic Models of Segregation. *Journal of Mathematical Sociology* 1 (1971), 143–186.

[22] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi. 2016. The spreading of misinformation online. *Proceedings of the National Academy of Sciences* 113 (2016), 554–559.

[23] Michela Del Vicario, Gianna Vivaldo, Alessandro Bessi, Fabiana Zollo, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. 2016. Echo Chambers: Emotional Contagion and Group Polarization on Facebook. *Scientific Reports* 6, 37825 (2016).

[24] Marcin Waniek, Tomasz P. Michalak, Michael J. Wooldridge, and Talal Rahwan. 2018. Hiding individuals and communities in a social network. *Nature Human Behaviour* 2 (2018), 139–147.