Selection and influence in cultural dynamics*

DAVID KEMPE

Department of Computer Science, University of Southern California,
Los Angeles CA 90089-0781, USA
(e-mail: `dkempe@usc.edu`*)*

JON KLEINBERG

Department of Computer Science, Cornell University, Ithaca NY 14853, USA
(e-mail: `kleinber@cs.cornell.edu`*)*

SIGAL OREN

Department of Computer Science, Ben-Gurion University of the Negev,
Beer Sheva 8410501, Israel
(e-mail: `sigalo@cs.huji.ac.il`*)*

ALEKSANDRS SLIVKINS

Microsoft Research, New York, NY 10011, USA
(e-mail: `slivkins@microsoft.com`*)*

Abstract

One of the fundamental principles driving diversity or homogeneity in domains such as cultural differentiation, political affiliation, and product adoption is the tension between two forces: influence (the tendency of people to become similar to others they interact with) and selection (the tendency to be affected most by the behavior of others who are already similar). Influence tends to promote homogeneity within a society, while selection frequently causes fragmentation. When both forces act simultaneously, it becomes an interesting question to analyze which societal outcomes should be expected.

To study this issue more formally, we analyze a natural stylized model built upon active lines of work in political opinion formation, cultural diversity, and language evolution. We assume that the population is partitioned into "types" according to some traits (such as language spoken or political affiliation). While all types of people interact with one another, only people with sufficiently similar types can possibly influence one another. The "similarity" is captured by a graph on types in which individuals of the same or adjacent types can influence one another. We achieve an essentially complete characterization of (stable) equilibrium outcomes and prove convergence from all starting states. We also consider generalizations of this model.

Keywords: *social networks, selection, influence, opinion formation*

1 Introduction

1.1 Selection and influence

Human societies exhibit many forms of cultural diversity—in the languages that are spoken, in the opinions and values that are held, and in many other dimensions. An

---

* A one-page abstract of this work has appeared in *ACM Conf. on Electronic Commerce*, 2013.

active body of research in the mathematical social sciences has developed models for reasoning about the origins of this diversity, and about how it evolves over time.

One of the fundamental principles driving cultural diversity is the tension between two forces: influence and selection. *Influence* refers to the tendency of people to become similar to those with whom they interact, whereas *selection* (or *choice homophily*, McPherson et al. (2001)) is the tendency of people to interact with those who are more similar to them, and/or to be more receptive to influence from those who are similar.[1]

Both of these forces lead toward outcomes in which people end up interacting with others like themselves, but in different ways: influence tends to promote homogeneity, as people shift their behaviors to become alike, while selection tends to promote fragmentation, in which a society can split into multiple groups that have less and less interaction with each other. Research that offers qualitative analyses for issues such as consensus-building, political polarization, or social stratification can often be interpreted through the lens of this influence-selection trade-off (Cohen, 1977; Kandel, 1978; McPherson et al., 2001). The trade-off between influence and selection, and the development of data analysis techniques to try separating the effects of the two, have been integral to understanding and promoting the adoption of products and behaviors in social networks (Anagnostopoulos et al., 2008; Aral et al., 2009; Bramoullé et al., 2009; LaFond & Neville, 2010; Shalizi & Thomas, 2011), an active line of work at the interface of computing, economics, and statistics.

When both influence and selection are operating at the same time, how should we reason about their combined effects? In particular, as Axelrod (1997, p. 203) asked:

If people who are similar to one another tend to become more alike in their beliefs, attitudes, and behavior when they interact, why do not all such differences eventually disappear?

Several lines of modeling work have approached this question, all starting from similar underlying motivations, but developing different mathematical formalisms.

1. Research on political opinions has studied populations in which each person holds an opinion. The opinion is represented by a number drawn from a bounded interval on the real line $\mathbb{R}^1$, or from a discrete set of points in an interval. (For example, the interval may represent the political spectrum from liberal to conservative.) Each person is influenced by the opinions of others who are sufficiently nearby on the interval, thus capturing the interplay between influence (people are shifting their opinions based on the opinions of others) and selection (people only pay attention to others whose opinions are sufficiently close) (Ben-Naim et al., 2003; Deffuant et al., 2000; Hegselmann & Krause, 2002).

2. Axelrod proposed a model of cultural diversity in which there are several *dimensions* of culture, and each person has a value associated with each dimension (e.g. a choice of language, religion, or political affiliation). Agents

---

[1] While *selection* may sometimes have causes other than similarity, such as attraction of the opposites or triadic closure, we focus on similarity-driven selection throughout this paper. We use the term *selection* rather than *homophily* because the latter is sometimes used to refer to the broader fact that people tend to be similar to their neighbors in a social network, regardless of the mechanism leading to this similarity.

are more likely to interact when they agree on more dimensions; when two people interact, one person randomly chooses a dimension in which they differ, and changes his value so that they now match in this dimension (Axelrod, 1997). For example, two people who have passions for similar sports and styles of food may end up having an easier time (and more opportunity for) associating, and hence an easier time influencing one another along another dimension such as religious beliefs. Again, the model represents an influence process in which the interactions are governed by selection based on (cultural) similarity. Axelrod's model has generated a large amount of subsequent work; see Castellano et al. (2009) for a survey.

3. Finally, Abrams and Strogatz exhibited some of the interesting effects that can occur even when there are only two types of people. They modeled a scenario in which people speak one of two languages. People mainly interact with speakers of their own language, but there is gradual "leakage" over time as speakers of one language may convert to become speakers of the other (Abrams & Strogatz, 2003). The Abrams–Strogatz model has also generated an active line of follow-up results, including explorations of its microfoundations through agent-based simulation (Stauffer et al., 2007) and analyses of the spatial effects and population density (Patriarca & Leppanen, 2004).

1.2 Commonalities among models

Although the models described above differ in many details, they have the same underlying structure: the population is divided into a set of *types* (the opinions, the cultural choices, the language spoken), and a person of any given type may be influenced to switch types, but only by others whose types are sufficiently similar. (In the case of the Abrams–Strogatz model, there is a preference for one's own type, but since there are only two types, all types can influence each other.) This process generates a "flow" as people migrate among different types, and we can ask questions about both dynamics (which outcomes the process will reach) and equilibria (which outcomes are self-sustaining, in the sense that the flows between types preserve the fraction of people who belong to each type). Following the language around Axelrod's work, we will refer to this type of process as representing the *cultural dynamics* of the population.

In addition to their similarities in structure, these cultural dynamics models also agree in their broad conclusions. In the first two models, the population gradually separates into distinct "islands" in the space of possible types; subsequently, no further interaction between the islands is possible. In the Abrams–Strogatz model, with just two types, the only outcomes that are stable under perturbations are the two extreme outcomes in which everyone ends up belonging to the same type. Typically, there is also an unstable equilibrium in which each language is spoken by a non-zero fraction of the population.

The most salient difference among the models is the structure that is imposed on the set of types. In each case, there is an undirected *influence graph* $\mathcal{T}$ on the set of types: when a person of type u interacts with a person of type v, the person of type u has the potential to switch to (or move towards) v *provided* that u and v are neighbors in $\mathcal{T}$ (i.e. provided that u and v are sufficiently similar according to the
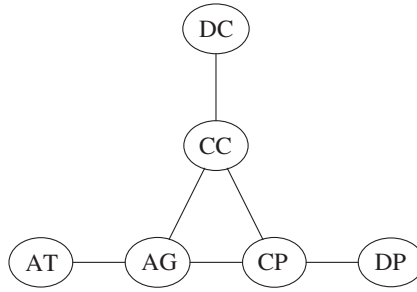
Fig. 1. A small subgraph of the "religion" graph, showing edges along which transitions may happen between agnostics, atheists, casual protestants, devout protestants, casual catholics, and devout catholics.

interpretation of the model). In the models of one-dimensional opinion dynamics on a discrete set, the graph $\mathcal{T}$ is the kth power of a path for some $k \geqslant 1$ (types are similar enough when they are within k steps on the path); in Axelrod's model, the graph $\mathcal{T}$ is the kth power of a (not necessarily binary) hypercube. The Abrams–Strogatz model shows that these kinds of processes can exhibit subtle behavior even on a two-node influence graph $\mathcal{T}$.

1.3 The present work: Cultural dynamics on an arbitrary influence graph

All of the prior results apply only to highly structured, symmetric graphs (essentially hypercubes and paths), whereas in some of the settings that the models seek to capture, the set of types can have a less orderly structure. As one simple example, consider a subgraph of the "religion" graph (depicted in Figure 1), with the following six types: agnostics (AG), atheists (AT), casual protestants (CP), devout protestants (DP), casual catholics (CC), and devout catholics (DC). Here, it is reasonable to assume that transitions happen between the casual versions of each belief, or between casual and devout versions of the same belief. In other words, the graph would consist of a triangle AG-CP-CC, and edges AT-AG, DP-CP, DC-CC.[2]

To ensure that insights derived from the analysis of a model (such as the ones for hypercubes and graphs) are not limited to those specific models, and to further understand the governing principles, it is desirable to understand the dynamics and equilibria of the process in more general graphs.

This is the problem we address in the present work, where we develop techniques for resolving some of the main questions on arbitrary graphs, under a clean and stylized model of interactions. For a natural formulation of cultural dynamics on an arbitrary influence graph (which we refer to as the *global model*, for reasons explained later), we prove convergence results and precisely characterize the set of all stable equilibria. We then consider generalizations of the global model, extending some of our convergence and stability results to these more general settings and posing several open questions.

---

[2] We may expect transitions between other states to happen, albeit with much smaller probability. We will discuss this issue more in Section 5.1.

1.4 The global model

We now describe the global model in more detail. Because the models from the earlier lines of work discussed above differ in many of their details, there is no meaningful way to simultaneously generalize all of them in a precise syntactic sense. Instead, our goal is to formulate a version of cultural dynamics that exhibits the same basic interplay of selection and influence—specifically, the idea that influence only happens among types that are "close together'—while allowing for an arbitrary graph on the set of types.

Let $\mathscr{T}$ be a graph on a finite set of types V of cardinality $n = |V|$; for each type $u \in V$, let $T_u \subseteq V$ denote the set of u's neighbors in $\mathscr{T}$. As is standard in many of the approaches to cultural dynamics, we model the population as a continuum[3]: at the start of the process, each type $u \in V$ has a non-negative population *mass* associated with it, corresponding to the fraction of the population that initially has this type. (Consider, for example, the fraction of the world's population that belongs to a certain religion or speaks a certain language.) Time evolves continuously[4] and $x_u(t)$ denotes the mass on type u at time t. The full state of the population at time t is thus given by the *mass vector* $x(t)$, the vector of values $x_u(t)$ for all $u \in V$.

We define a continuous-time dynamical system in which the direction in which the populations move is determined in terms of the mass vector $x(t)$. The dynamical system is motivated by imagining that each person chooses a random other person to interact with. Selection effects are captured in two ways by the model: first, people are more likely to interact with their type; second, they only have the potential to be influenced when they interact with an individual of their own or a neighboring type. Specifically, each person is α times more likely to choose an interaction partner of their own type than someone of a different type, for a parameter $\alpha \geqslant 1$. When a person of type u chooses to interact with a person of type v, such that $v \in T_u$, with probability p, he will switch to type v, where $p \in (0, 1]$ is a fixed parameter. To express this dynamic numerically, we let $\mathscr{M}_u(t) = \alpha x_u(t) + \sum_{v \in V \setminus \{u\}} x_v(t)$. A person of type u chooses to interact with his own type with probability $\alpha x_u(t)/\mathscr{M}_u(t)$, and with any other type $v \neq u$ with probability $x_v(t)/\mathscr{M}_u(t)$. Thus, the fraction of the *entire population* which is moving from u to v is $p \cdot x_u(t)x_v(t)/\mathscr{M}_u(t)$. At the same time as this mass of $p \cdot x_u(t)x_v(t)/\mathscr{M}_u(t)$ is moving from u to v, a mass of $p \cdot x_v(t)x_u(t)/\mathscr{M}_v(t)$ is moving from v to u. These movements partially cancel each other out, and motivate the following definition of the (directed) *flow* on the edge $(v, u) \in \mathscr{T}$:

$$f_{v \to u}(t) = p \cdot x_v(t)\, x_u(t) \left(\frac{1}{\mathscr{M}_v(t)} - \frac{1}{\mathscr{M}_u(t)} \right). \tag{1}$$

The change in mass at a node u can then be written as

$$\dot{x}_u(t) = \sum_{v \in T_u} f_{v \to u}(t) \; = \; p \cdot x_u(t) \sum_{v \in T_u} x_v(t) \left(\frac{1}{\mathscr{M}_v(t)} - \frac{1}{\mathscr{M}_u(t)} \right). \tag{2}$$

Notice that because the system is characterized by a system of differential equations and that for all u the derivative of $x_u(t)$ is finite (by Equation (2)), we obtain that $x_u(t)$ is continuous for all u.

---

[3] This and other modeling choices are discussed in more detail in Section 5.1.
[4] Again, refer to Section 5.1 for a discussion.

It is natural to think of the parameter p as generally being very small, since most interactions between people do not lead to a change of type. However, as it turns out, the value of p does not have a major qualitative effect on our results. This is not surprising, since introducing $p < 1$ (as opposed to $p = 1$) just slows down the flow between any two types by a factor of $1/p$. We include p in the model in order to capture the range of possible speeds at which transitions can happen. For example, if the types in our model correspond to dialects of a language, we can choose a small p (since the probability that a person changes his dialect is very small). However, if instead the types represent opinions in the period before an election, people may switch much more rapidly, and a larger p is appropriate.

1.5 Convergence, equilibria, and stability in the global model

Our first result is that for any influence graph $\mathscr{T}$ and any initial mass vector x the system converges to a limit mass vector x^*. We prove this by establishing a system of invariants on the population masses over time, capturing a certain "rich-get-richer" property of the process—essentially, that the types of large mass will tend to grow at the expense of the types of small mass.

We next consider the equilibria of this model: we say that a mass vector x is an equilibrium if it remains unchanged after one application of the update rule. It is easy to construct examples of equilibria that are not stable, in the sense that an arbitrarily small perturbation of the masses x_u^* can—after further applications of the update rule—push the masses far away from the equilibrium. Such equilibria are less natural as predicted outcomes of the cultural dynamics being modeled, since the population would be unlikely to hold its position near this equilibrium.

To make this statement precise, we use the notion of *Lyapunov stability*. We say that an equilibrium x^* is *Lyapunov stable* if given any $\varepsilon > 0$, there exists a $\delta > 0$ such that if $||x(t_0) - x^*||_1 < \delta$, then $||x(t) - x^*||_1 < \varepsilon$, for all $t \geqslant t_0$.[5] For simplicity, we use the L_1 norm throughout. Since our vectors have finite dimensionality, the different L_p norms only differ by constant factors, so all stability results apply to other norms by scaling ε, δ appropriately.

We prove that x^* is a Lyapunov-stable equilibrium if and only if the set of active types $A(x^*) = \{u : x_u^* > 0\}$ is an independent set in the influence graph $\mathscr{T}$. The proof is based on the rich-get-richer properties of the process; these properties are used to show that after a sufficiently small perturbation to the population masses, the amount by which any type with positive mass can grow is bounded.

1.6 Interpretations of the basic results

The basic results discussed above establish a precise sense in which the natural equilibria tend to break the population into non-interacting islands.

---

[5] One could ask about stronger notions of stability, in particular, *asymptotic stability*, which requires that there exists a $\delta_1 > 0$ such that if $||x(t_0) - x^*||_1 < \delta_1$, then $x(t) \to x^*$ as $t \to \infty$. Asymptotic stability is not a useful definition for our purposes; for example, if the underlying influence graph $\mathscr{T}$ has no edges, then any assignment of population masses is an equilibrium, but none are asymptotically stable, since there is no way for a small perturbation to converge back to the original state. On the other hand, all equilibria are Lyapunov-stable in this simple example.

In addition to offering a qualitative statement about fragmentation of opinions under the proposed stylized model, the results also suggest a way of reasoning about the phenomenon by which opinions on different issues tend to become aligned, with an individual's views on one issue providing evidence for his or her views on another (Poole & Rosenthal, 1991; Spector, 2000). To take a concrete example that already appears on the two-dimensional hypercube (i.e. the 4-node cycle), consider a setting in which each individual has either a liberal or conservative view on fiscal issues and either a liberal or conservative view on social issues. If we assume that people only influence each other when they agree on at least one of these two categories of issues, then the graph on the set of types is a 4-node cycle. Since our results on Lyapunov-stable equilibria indicate that independent sets are favored as outcomes, we can interpret the conclusion in this example as predicting that under the proposed model, either the whole population will converge on a single node (representing a uniform choice of views), or on an independent set of two nodes, in which case an individual's opinion on fiscal issues has become correlated with his or her opinion on social issues.

It is also instructive to compare our results to the main result of Abrams & Strogatz (2003) discussed above. Recall that they consider the influence graph $\mathscr{T} = K_2$ (two connected nodes), and they find that the two stable equilibria are the outcomes in which all the population mass is gathered at a single node. The family of dynamical systems they consider strictly subsumes ours in the special case of a two-node graph, but for the specific system we study, our results imply that their basic finding extends to arbitrary graphs: in any graph, the Lyapunov-stable equilibria correspond to the non-empty independent sets, just as Abrams and Strogatz showed for the two-node graph K_2.

1.7 A generalization: Limiting both interaction and influence

We now discuss a natural generalization of the model that is significantly more challenging to analyze. In the global model, the members of type u can interact with members of *all* other types, even though they are influenced only by the types in T_u. However, there are settings in which it is more natural to assume that the members of a type only ever *interact* with members of a subset of the other types; for example, this may be a reasonable assumption when types represent different languages. This is somewhat similar to the approach Centola et al. (2007) in studying a variant of the Axelrod model. Under this variant, there is a social network among the agents; whenever two agents become so different that they cannot influence one another any more, the tie between them is broken, and new ties are formed. Centola et al. (2007) use simulations to show that multi-cultural equilibria form readily and stably under this model.

To capture the idea that some types may simply be too different to interact, we now assume that there are two potentially distinct graphs on the set of types V: the influence graph $\mathscr{T}$ (as before), as well as an undirected *interaction graph* $\mathscr{S}$, where $\mathscr{T}$ is a subgraph of $\mathscr{S}$. Rather than interacting with a person chosen from the full population, a member of type u selects an interaction partner from the set S_u of u's neighbors in $\mathscr{S}$. It is straightforward to write the new update rule for this more general dynamical system, by summing over types in S_u instead of $V \setminus \{u\}$.

Specifically, we can now define

$$\mathcal{M}_u(t) = \alpha x_u(t) + \sum_{v \in S_u} x_v(t). \tag{3}$$

With this new definition of $\mathcal{M}_u(t)$, the definitions of flows and updated masses at nodes from Equations (1), (2) stay exactly the same. Hence, the $\mathcal{M}_u(t)$ terms emerge as crucial quantities that determine the direction of the flow; for that reason, we will call $\mathcal{M}_u(t)$ the *interaction mass* of node u at time t.

The global model is simply the special case in which the interaction graph $\mathcal{S}$ is the complete graph. The name *global model* emphasizes that each type interacts "globally," with all other types.[6]

The behavior of this general model is significantly more complex than the behavior of the global model; for instance, for arbitrary $\mathcal{S}$ and $\mathcal{T}$, it is not even clear whether the process will always converge. Intuitively, much of the difficulty comes from the fact that when we consider two neighboring types u and v, the sets of types that they are interacting with, S_u and S_v, can be quite different, whereas in the global model they are both the full set V. Among other things, this can lead to violations of the rich-get-richer property that was so useful for reasoning about the dynamics of the global model.

For the general model, we first establish a necessary condition for equilibria, as well as sufficient conditions for convergence and stability. We then focus further on the special case in which $\mathcal{S} = \mathcal{T}$. This is in a sense the opposite extreme from the global model; instead of making $\mathcal{S}$ as large as it can be, we make it as small as possible subject to the constraint that it contains $\mathcal{T}$ as a subgraph. Accordingly, we refer to the case $\mathcal{S} = \mathcal{T}$ as the *local model*. There are many interesting open questions surrounding the behavior of the local model; we make progress on these through initial convergence results and the identification of a large class of equilibria that are Lyapunov-stable for all $\alpha > 1$: non-empty independent sets for which all nodes in the set are at a mutual distance of at least three. In fact, this is an "if and only if" characterization for an important class of influence graphs: those whose connected components are trees or, more generally, bipartite graphs.

An interesting observation is that the local and global models can have genuinely different behaviors starting from the same initial conditions: Figure 2 shows an example of an initial mass distribution on the 3-node path for which the global model converges to an outcome in which the mass is divided evenly between the two endpoints, while the local model converges to the outcome in which all the mass is on the middle node.

At a higher level, formalizing the distinction between interaction ($\mathcal{S}$) and influence ($\mathcal{T}$) is a potentially promising activity more broadly, particularly in light of the considerable recent interest in the effects of information filtering on the political process. (See Pariser (2011); Sunstein (2009) for popular media accounts, and Bakshy et al. (2012) for recent experimental research.) The concern expressed in

---

[6] There are clearly many other potential generalizations which could incorporate notions of non-uniform interaction, including different interaction strengths between different pairs of types. Such extensions would lead to interesting questions as well. In the present work, we focus on the generalization with two unweighted graphs $\mathcal{S}$ and $\mathcal{T}$ because it captures in a direct way some of the additional complexity that is introduced by simultaneously modeling limited interaction and influence.

(a) Initial fractions.

(b) Global model ($\alpha = 2$).
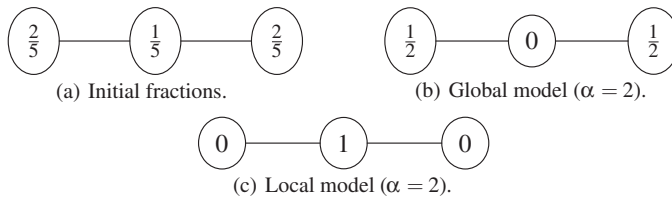
(c) Local model ($\alpha = 2$).

Fig. 2. An instance in which different models predict convergence to different equilibria. The global model predicts an outcome in which two non-interacting types survive (polarization), whereas the local model predicts that only a single types survives (consensus).

all these lines of work is that personalization on the Internet makes it possible to sharply restrict the diversity of information one sees, and thus risks accentuating the degree of polarization and fragmentation in political discourse—essentially, the risk is that people will only ever be exposed to those who already agree with them, making any kind of consensus almost impossible to achieve.

In this context, our general model also brings into the discussion the interesting contrast between interaction and influence. Personal filtering of information by Internet users can restrict the set of people they interact with (affecting the sets S_u), and it can also, separately, restrict the set of people who may be able to influence them (affecting the sets T_u). These two different effects are often bundled together in discussions of information filtering; it will be interesting to see whether treating them as genuinely distinct can shed additional light on this set of issues.

1.8 Additional related work

Steglich et al. (2010) (see also Snijders et al. (2007)) provide a general model of social networks that combines selection and influence. While we focus on deriving structural properties of a network, these papers pursue a different goal: statistically valid inference of network parameters from real-life observations. More broadly, inferring latent properties of a social network from observations has been an active line of work. Some of the notable directions in this work, aside from the one taken in Steglich et al. (2010); Snijders et al. (2007), include latent "social space" reconstruction (e.g. Handcock et al. (2007); Hoff et al. (2002)) and community detection (e.g. see Brandes & Erlebach (2005); Schaeffer (2007); Fortunato (2010)).

2 Observations on the general model

In this section, we develop several observations that apply to the fully general model with an arbitrary interaction graph $\mathscr{S}$. In the subsequent sections, we utilize these observations to analyze the global model (where $\mathscr{S}$ is the complete graph) and the local model (where $\mathscr{S} = \mathscr{T}$).

We say that a node u is *active* at time t if $x_u(t) > 0$, and *inactive* if $x_u(t) = 0$. We occasionally refer to a node u as *x-active* if it is active in x and *x-inactive* otherwise. The set of all active nodes under x is denoted by $A(x)$. Much of our analysis concerns the structure of the subgraph $\mathscr{T}_{\mathtt{act}}(x)$ of the influence graph $\mathscr{T}$

induced by the active nodes $A(x)$. We begin by characterizing when mass vectors are in equilibrium.

Proposition 2.1

A vector x^* is an equilibrium if and only if each connected component C of $\mathcal{T}_{\mathrm{act}}(x^*)$ has the property that all nodes $u \in C$ have the same interaction mass $\mathcal{M}_u$.

Proof

x^* is at equilibrium if and only if the flow on all edges is 0. In turn, from Equation (1), we see that the flow on the edge (u, v) is 0 if and only if at least one of the following two conditions holds: (1) $\mathcal{M}_u = \mathcal{M}_v$, (2) $x_u^* \cdot x_v^* = 0$.

If x^* satisfies the assumptions, then each edge (u, v) is either inside a component (and thus $\mathcal{M}_u = \mathcal{M}_v$) or has at least one inactive endpoint (and thus $x_u^* \cdot x_v^* = 0$). Conversely, if x^* is an equilibrium, each edge satisfies (1) or (2). When u, v lie in the same component C, there is a path between them in C, and along that path, (1) must hold for all edges, so u and v must have the same interaction mass. $\square$

The following useful lemma relates convergence and the change in directions of flows:

Lemma 2.2

If there exists a time t_0 such that the flows do not change direction after time t_0, then the system converges.

Proof

Let G be the directed graph obtained by directing each edge (u, v) of $\mathcal{T}$ according to the direction of the corresponding flow $f_{u \to v}(t_0)$. By the assumption, these directions stay constant after time t_0. As flow always goes from types with smaller interaction mass to types with larger interaction mass, G must be acyclic. Let $v_1, v_2, \ldots, v_n$ be a topological sorting of the graph, so that all directed edges of G are of the form $(v_i, v_j), i < j$.

We define $X_k(t) = \sum_{i=1}^{k} x_{v_i}(t)$ to be the total mass at time $t \geq t_0$ on the k first nodes in the topological sorting. Because the total mass in the system is constant, and all flow goes from nodes with lower indices to nodes with higher indices, each of the $X_k(t)$ must be non-increasing as a function of t. Since they are also lower-bounded by 0, each $X_k(t)$ must converge to some value Z_k as $t \to \infty$. Therefore, each $x_{v_i}(t)$ converges to $Z_i - Z_{i-1}$ as $t \to \infty$. $\square$

Recall that we are interested in characterizing Lyapunov-stable equilibria. We next provide a sufficient condition.

Proposition 2.3

An equilibrium x^* is Lyapunov-stable if it satisfies the following two properties:

1. The active nodes form an independent set in the influence graph $\mathcal{T}$.
2. The interaction mass of every active node is strictly greater than the interaction mass of each of its inactive neighbors in the influence graph $\mathcal{T}$.

Proof

Let x^* be an equilibrium for which both properties hold. A node u is called x^*-*active* if it is active in x^* and x^*-*inactive* otherwise. Let A be the set of all x^*-active nodes, and let $\mathcal{M}_u^*$ denote the interaction mass of node u with respect to x^*. Define

$$\delta = \tfrac{1}{2\alpha+1} \min_{u \in A,\, v \notin A,\, (u,v) \in \mathcal{T}} (\mathcal{M}_u^* - \mathcal{M}_v^*) > 0 \qquad (4)$$

by the second property of x^*. To show stability, we prove that whenever $||x^* - x(t_0)||_1 \leqslant \delta$, the system will satisfy $||x^* - x(t)||_1 \leqslant \delta$ for all times $t \geqslant t_0$.

The key step of the proof is to establish that for each node $u \in A$, the mass $x_u(t)$ is non-decreasing over time, i.e. that $\dot{x}_u(t) \geqslant 0$ for all $t \geqslant t_0$. The initial condition implies that $x_u(t_0) \geqslant x_u^* - \delta$ for all $u \in A$, $\sum_{u \in A} x_u(t_0) \geqslant \sum_{u \in A} x_u^* - \delta$ and $\sum_{v \notin A} x_v(t_0) \leqslant \delta$. We will show that these invariants are maintained for all $t \geqslant t_0$.

Consider any time $t \geqslant t_0$ and edge (u, v) with $u \in A$ and $v \notin A$. The invariants imply that $\mathcal{M}_u(t) \geqslant \mathcal{M}_u^* - \alpha\delta$ and $\mathcal{M}_v(t) \leqslant \mathcal{M}_v^* + \alpha\delta$, so we obtain that

$$\mathcal{M}_u(t) - \mathcal{M}_v(t) \geqslant (\mathcal{M}_u^* - \mathcal{M}_v^*) - 2\alpha\delta \stackrel{(4)}{\geqslant} (2\alpha + 1)\delta - (2\alpha)\delta > 0.$$

In particular, this implies that $f_{v \to u}(t) \geqslant 0$; because this holds for all $v \notin A$, we have established that $\dot{x}_u(t) \geqslant 0$. By summing overall nodes $u \in A$, we have shown that the invariant continues to hold.

Finally, because each of the $x_u(t), u \in A$ is non-decreasing, mass can only move among x^*-inactive nodes, or from x^*-inactive nodes to x^*-active ones. Therefore, $\sum_{u \in A} |x_u(t) - x_u(t_0)| = \sum_{u \notin A} x_u(t_0) - \sum_{u \notin A} x_u(t)$. Thus,

$$\begin{aligned}
||x(t) - x^*||_1 &\leqslant \sum_{u \in A} |x_u(t) - x_u(t_0)| + \sum_{u \in A} |x_u(t_0) - x_u^*| + \sum_{u \notin A} x_u(t) \\
&= \sum_{u \in A} |x_u(t_0) - x_u^*| + \sum_{u \notin A} x_u(t_0) = ||x(t_0) - x^*||_1 \leqslant \delta,
\end{aligned}$$

so the system is Lyapunov-stable. $\qquad\square$

3 The global model

In this section, we analyze the global model. The definition of the general model states that flows are always directed from nodes with smaller interaction mass to nodes with larger interaction mass. For the global model, this property is simplified significantly: flow is always directed from types with smaller mass to types with larger mass. This property lets us achieve an almost complete understanding of the global model. We show that for this model, the system always converges, and we present a complete characterization of which equilibria are Lyapunov-stable. First, we characterize equilibria by applying Proposition 2.1 to the global model.

Corollary 3.1

Under the global model with $\alpha > 1$, the system is at equilibrium x^* if and only if the following holds: for every connected component C of $\mathcal{T}_{\mathsf{act}}(x^*)$, all nodes $u \in C$ have the same mass.

Proof
Proposition 2.1 guarantees that x^* is at equilibrium if and only if for each edge $(u, u') \in \mathcal{T}_{\text{act}}(x^*)$: $\alpha x_u^* + \sum_{v \in S_u} x_v^* = \alpha x_{u'}^* + \sum_{v \in S_{u'}} x_v^*$. In the global model, for any node u, the set S_u consists of all types but u itself, implying that the sum cancels out, and we obtain $(\alpha - 1)x_u^* = (\alpha - 1)x_{u'}^*$. For $\alpha > 1$, this implies $x_u^* = x_{u'}^*$. $\qquad\square$

We next show that the system always converges; the proof relies on the key invariant that for any $1 \leqslant k \leqslant n$, the total mass of the k smallest types never increases over time. More formally, we define the following quantities:

Definition 3.2
Let $y_1(t) \leqslant y_2(t) \leqslant \ldots \leqslant y_n(t)$ be the node masses sorted in non-decreasing order. Define

$$Y_k(t) = \sum_{i \leqslant k} y_i(t) = \min_{R : |R| = k} \sum_{v \in R} x_v(t) \tag{5}$$

to be the sum of the masses of the k smallest nodes at time t.

The following lemma formally captures the notion that the rich get richer in the global model.

Lemma 3.3
For every k, the function $Y_k(t)$ is non-increasing in t, i.e. $\dot{Y}_k(t) \leqslant 0$.

Proof
Let t, k be arbitrary. Consider any set S of k nodes achieving the minimum in Equation (5) at time t; notice that there could be multiple such sets S. Consider any $u \in S, v \notin S$; by definition of S, we have that $x_u(t) \leqslant x_v(t)$, and hence $f_{u \to v}(t) \geqslant 0$. Because this holds for all such edges (u, v), we obtain that the total weight on nodes of S cannot increase. As this holds for all candidate sets S, we get that $\dot{Y}_k(t) \leqslant 0$. $\qquad\square$

Theorem 3.4
Under the global model, the system converges for any influence graph and any starting mass vector $x(0)$.

Proof
By Lemma 3.3, each function $Y_j(t)$ is non-increasing in t. As all masses are non-negative, the $Y_j(t)$ are also bounded below by 0. Hence, each function $Y_j(t)$ must converge to some value Z_j. Thus, each function $y_j(t)$ must converge to $Z_j - Z_{j-1} =: z_j$. It remains to show that this also implies convergence of $x(t)$.

Let $\delta > 0$ be at most the smallest difference between any two distinct z_j, i.e. $\delta \leqslant \min_{i, j : z_i \neq z_j} |z_i - z_j|$. Let t_0 be large enough that $|y_i(t) - z_i| < \frac{\delta}{3}$ for all i and $t \geqslant t_0$.

We will show that the only cases in which there could be nodes v and times $t' > t \geqslant t_0$ such that $x_v(t) = y_j(t)$ and $x_v(t') = y_{j'}(t')$ is to have $z_j = z_{j'}$. If not, then let $\hat{t}$ be such that $|x_v(t) - z_j| < \delta/3$ for $t < \hat{t}$ arbitrarily close to $\hat{t}$, and $|x_v(t) - z_{j'}| < \delta/3$ for $t' > \hat{t}$ arbitrarily close to $\hat{t}$. Because $|z_{j'} - z_j| \geqslant \delta$, this implies that $x_v(t)$ must be discontinuous at $t = \hat{t}$, which it cannot be. $\qquad\square$

3.1 Characterization of Lyapunov-stable equilibria

For the global model, the properties required for Proposition 2.3 hold for any independent set, since the interaction mass of active types is always greater than the interaction mass of inactive types. Therefore, any equilibrium in which the set of active nodes is independent is Lyapunov-stable. To complete the characterization, we show that the converse is also true.

Theorem 3.5
In the global model with $\alpha > 1$, an equilibrium x^* is Lyapunov-stable if and only if the active nodes form an independent set.

Proof
It remains to prove the "only if" direction. Assume that the active nodes in an equilibrium x^* do not form an independent set. We will prove that x^* is not Lyapunov-stable.

Let C be a connected component of size $|C| \geqslant 2$ in $\mathscr{T}_{\text{act}}(x^*)$. By the assumption that the active nodes in x^* do not form an independent set, such a connected component exists. Notice that each component of $\mathscr{T}_{\text{act}}(x^*)$ evolves in isolation, so we can focus on only C for the rest of the proof. Therefore, by Corollary 3.1, $x_v^* = \mu$ for all $v \in C$, for some value μ.

Let $u, v \in C$ be two arbitrary nodes, and $\delta > 0$ be arbitrarily small. Consider the following perturbation: $x_u = x_u^* + \delta, x_v = x_v^* - \delta$, and $x_w = x_w^*$ for all $w \neq u, v$. By Theorem 3.4, the system, starting from the perturbed vector x, will converge to some new equilibrium y. By Lemma 3.3, the smallest mass of any node in C will always be at most $\mu - \delta$ during the process. All y-active nodes must have the same mass; therefore, if all nodes were active in y, they would all have to have mass at most $\mu - \delta$, which would imply that mass has disappeared from C, a contradiction. Hence, at least one node of C must end up inactive in y. In particular, this means that $||x^* - y||_1$ is not bounded in terms of δ, and x^* is not Lyapunov-stable. □

4 The local model

In the previous section, we have given essentially complete characterizations of convergence and stability of equilibria under the global model, in which all types have the potential to interact, even though only certain pairs of types can influence each other (according to the graph $\mathscr{T}$).

We now consider the local model, which is at the other extreme of our general family: here, the interaction graph $\mathscr{S}$ is the same as the influence graph $\mathscr{T}$; hence, interactions occur only between individuals who also have the potential to influence each other. (We will generally denote this underlying graph by $\mathscr{T}$, with the understanding that $\mathscr{S} = \mathscr{T}$.) We find that the problems of convergence and stability are much more challenging in this case. For the global model, we were able to extract very useful organizing structures in the dynamical system that gave us a natural progress measure toward convergence. But as is well known, in general, a dynamical system on even a small number of variables may have convergence properties that are extremely difficult to analyze or express. For example, not only does Lemma 3.3 not hold for the node masses; a reformulation for *interaction masses* does not hold, either.

Given the complex behavior of the update rules for the local model, we find that the convergence and stability questions are already difficult on graphs $\mathscr{T}$ with a small number of nodes, and we focus our results here on such cases. (Of course, based on the motivating premise of the model, even systems with a small number of variables are frequently natural, corresponding to selection and influence dynamics in societies with, for example, a small number of languages, a small number of political parties, or a small number of dominant religions or cultures.)

We begin by considering the case $\alpha > 1$ and first prove the following theorem:

Theorem 4.1
Under the local model, if the influence graph is a 3-path, then the system converges from any starting state.

The full proof is provided in Appendix A, but we provide a brief outline here. The subtle difficulty arises due to the fact that the flow between two types u and v does not necessarily go in the same direction at all times, but instead may change its direction. To keep track of the changes in direction, we define a *configuration* of the system to be a labeling of all edges (u, v) in $\mathscr{T}$ by the direction along which flow is traveling (i.e. whether from u to v or from v to u). In the case of a 3-node path, there are four possible configurations. We study transitions among the configurations as the system evolves over time; we show that each configuration is either a *sink*, which cannot transition to any other configuration, or it has the property that any change in the direction of an edge leads to a sink configuration. This ensures that there can be at most one change in the direction of flow as the system evolves; hence, there is a time t_0 such that for any $t > t_0$, no flow changes its direction. After this point, Lemma 2.2 guarantees that the system converges. For the case $\alpha \geqslant 2$, we show this fact only for the 3-path; for $\alpha < 2$, we establish a more general result, showing the same fact for arbitrary star graphs.

For $\alpha = 1$, we are able to prove convergence if the active subgraph is a path of $n \leqslant 5$ nodes. The proof requires different techniques than the ones we use for $\alpha > 1$: for paths of more than 3 nodes, flows on edges can change their direction infinitely often. The proof is provided in Appendix B.

4.1 Characterization of universally stable equilibria

Next, we turn our attention to Lyapunov-stable equilibria. We focus on a very strong notion of stability: stability of an equilibrium x simultaneously for all $\alpha > 1$.[7] Formally, we call a mass vector x a *universally stable equilibrium* if x is a Lyapunov-stable equilibrium for every $\alpha > 1$. Our goal here is to investigate which equilibria are universally stable. Such equilibria are robust to (a very idealized notion of) a change in the environment, as expressed by varying α.

Our main result for universally stable equilibria is a complete characterization for influence graphs that are forests, and more generally for influence graphs whose connected components are bipartite graphs.

---

[7] Contrast this with the notion of stability used in the previous section—there, we characterized Lyapunov-stable equilibria for any given fixed α.

Theorem 4.2
Assume that all connected components of the influence graph $\mathscr{T}$ are bipartite graphs. Then, a mass vector x^* is a universally stable equilibrium under the local model if and only if the distance between any two active nodes in x^* is at least 3.

The proof of Theorem 4.2 consists of several sub-results, all of which hold for arbitrary influence graphs, and imply the desired characterization under the assumption in the theorem. It is worth noting that these sub-results constitute significant progress towards understanding the structure of universally stable equilibria for arbitrary influence graphs, as we discuss later. For brevity, if x is a mass vector such that the distance between any two active nodes is at least 3, we will say that x is *3-separated*.

The first proposition proves the "if" direction of Theorem 4.2. Its proof follows from our analysis in Section 2.

Proposition 4.3
Under the local model with $\alpha > 1$, any 3-separated mass vector x^* is a Lyapunov-stable equilibrium.

Proof
x^* is an equilibrium by Proposition 2.1 since its active nodes form an independent set. By Proposition 2.3, an equilibrium whose active nodes form an independent set is Lyapunov-stable if the interaction mass of each active node is greater than the interaction mass of each of its neighbors. For $\alpha > 1$, this property holds when each inactive node has at most one active neighbor. In turn, this holds if and only if the distance between every two active nodes is at least 3; thus, all such equilibria are Lyapunov-stable for every $\alpha > 1$. □

The "only if" direction of Theorem 4.2 is more complicated to prove. First, we show that if the active nodes of an equilibrium do form an independent set, then being 3-separated is necessary to ensure universal stability.

Proposition 4.4
Let x^* be a universally stable equilibrium, and assume that the active nodes under x^* form an independent set. Then, x^* is 3-separated.

Proof
For the sake of contradiction, suppose that x^* is not 3-separated. Then, there exists an inactive node u with at least two active neighbors. Let A_u be the set of all active neighbors of node u. We use the following notation:

$$s = \sum_{v \in A_u} x_v^*; \quad \eta = \min_{v \in A_u} x_v^*; \quad \mu = \max_{v \in A_u} x_v^*.$$

Define $\alpha = 1 + \eta^2$. We will show that x^* is not Lyapunov-stable for this α.

To prove instability, consider a perturbation x which coincides with x^* on all nodes not in $\{u\} \cup A_u$, and satisfies

$$\begin{cases} x_v \leqslant x_v^* & \text{for all } v \in A_u \\ x_u = \delta & \text{for some } \delta \in (0, s - \mu - \eta^2) \\ x_u + \sum_{v \in A_u} x_v = \sum_{v \in A_u} x_v^* = s. \end{cases} \tag{6}$$

(Note that $s - \mu - \eta^2 > 0$ because A_u consists of at least two nodes.)

Under such an x, the interaction mass of node u is

$$\mathcal{M}_u \; = \; \alpha x_u + \sum_{v \in A_u} x_v \; = \; \eta^2 \cdot x_u + s,$$

while the interaction mass of any node $v \in A_u$ is

$$\mathcal{M}_v \; = \; \alpha x_v + x_u \; \leqslant \; \alpha x_v^* + x_u \; \leqslant \; (1 + \eta^2)\mu + x_u.$$

Since $x_u < s - \mu - \eta^2$, we have that $\mathcal{M}_v < (1 + \eta^2)\mu + s - \mu - \eta^2 < s$, and hence, $\mathcal{M}_u > \mathcal{M}_v$. Thus, under this perturbation, mass starts flowing from all nodes $v \in A_u$ to u, and this continues until $x_u \geqslant s - \mu - \eta^2$. Consequently, the system cannot reach any equilibrium with $x_u < s - \mu - \eta^2$; in particular, it cannot reach any equilibrium with $\|x^* - x\|_1 < s - \mu - \eta^2$. Since this holds for arbitrarily small δ, and $s - \mu - \eta^2 > 0$ is a constant independent of δ, we conclude that x^* is not Lyapunov-stable for this α. $\qquad\square$

With Proposition 4.4 in place, all that remains to complete the proof of the "only if" direction of Theorem 4.2 is to ensure that the active nodes in any universally stable equilibrium x^* of a bipartite graph form an independent set, i.e. that each connected component C of $\mathcal{T}_{\text{act}}(x^*)$ consists of a single node. This is implied by Lemma 4.5, which shows in general that if x^* is a universally stable equilibrium, then all the non-trivial connected components of the subgraph of its active nodes are *not* bipartite graphs. This completes the proof of Theorem 4.2, as any connected subgraph of a bipartite graph is a bipartite graph itself.

An additional benefit of Lemma 4.5 is that it applies to arbitrary influence graphs, and significantly limits the topologies a connected component of $\mathcal{T}_{\text{act}}(x^*)$ can have for a universally stable equilibrium x^*. To state this lemma in the most general form, we define a class of regular graphs which in particular subsumes all bipartite graphs, all cliques, and all cycles whose length is a multiple of 3. We say that a d-regular graph is *locally balanced* if its vertices can be partitioned into k disjoint sets $V_1, V_2, \ldots, V_k$ such that each vertex $v \in V_i$ has exactly $d/(k-1)$ edges to each of the sets $V_j, j \neq i$.

Lemma 4.5
Let x^* be a universally stable equilibrium and C a non-trivial connected component of its active subgraph $\mathcal{T}_{\text{act}}(x^*)$. Then:

    a. C is a regular graph, and x^* is uniform on C (i.e. $x_u^* = x_v^*$ for all $u, v \in C$).
    b. C is *not* a bipartite graph, and, more generally, C is not locally balanced.

Proof
We begin by proving part (a). Let $u, v \in C$ be a pair of *adjacent* nodes. The equilibrium conditions for $\alpha = 2$ imply that $2x_v^* + \sum_{w \in T_v} x_w^* = 2x_u^* + \sum_{w \in T_u} x_w^*$, and the ones for $\alpha = 3$ that $3x_v^* + \sum_{w \in T_v} x_w^* = 3x_u^* + \sum_{w \in T_u} x_w^*$. Subtracting the first equation from the second shows that $x_v^* = x_u^*$. Because C is a connected component, applying this argument along all edges in C proves that all nodes in C must have the same mass μ.

The interaction mass of node v with $\alpha = 2$ is therefore $\mathcal{M}_v = \mu \cdot (|T_v \cap C| + 1)$. Considering again a pair u, v of adjacent nodes, the equilibrium condition $\mathcal{M}_u = \mathcal{M}_v$

implies that $|T_u \cap C| = |T_v \cap C|$. Again by connectivity of C, this implies that all nodes in C have the same degree, so C is regular.

Next, we prove part (b). Because x^* is universally stable, part (a) implies that C is d-regular for some $d \geqslant 1$, and $x_u^* = \mu$ (for some μ) for all $u \in C$. Assume for contradiction that C is locally balanced, and let $V_1, \ldots, V_k$ be the k partitions of C. Because[8] $\mathcal{T}[V_i \cup V_j]$ is a $d/(k-1)$-regular bipartite graph for each pair $i \neq j$, all partitions V_i must have the same size $s = |C|/k$.

Set $\alpha = d + 1$, and let $\delta > 0$ be arbitrary. Consider perturbed vectors of the following form: $x_v = x_v^* + \frac{1}{s} \cdot \delta$ for every $v \in V_1$ and $x_u = x_u - \frac{1}{s(k-1)} \cdot \delta$ for every $u \notin V_1$. (That is, a total mass of δ is removed uniformly from nodes not in V_1, and added uniformly over the nodes in V_1.)

In moving from x^* to x, the interaction mass of each node $v \in V_1$ changes by $\alpha \cdot \frac{1}{s} \cdot \delta - d \cdot \frac{1}{s(k-1)} \cdot \delta > 0$, while the interaction mass of each node $u \notin V_1$ changes by

$$-\alpha \cdot \frac{1}{s(k-1)} \cdot \delta - \frac{d(k-2)}{k-1} \cdot \frac{1}{s(k-1)} \cdot \delta + \frac{d}{k-1} \cdot \frac{1}{s} \cdot \delta$$
$$= \left(-(d+1) - \frac{d(k-2)}{k-1} + d \right) \cdot \frac{1}{s(k-1)} \cdot \delta \ < \ 0.$$

Thus, for any such vector $x(t) = x$, all flows are directed from nodes not in V_1 to nodes in V_1. Furthermore, by symmetry of the original vector x^* and the perturbation, the mass vectors $x(t')$ for $t' > t$ will be of the same form, for a different $\delta' > \delta$. Thus, the same argument will apply at all times. Hence, the direction of flows never changes, and Lemma 2.2 guarantees that the system converges. Since the interaction mass of all nodes in V_1 is only increasing, and the interaction mass of all nodes not in V_1 is only decreasing, the only equilibrium y the system can converge to is one in which all nodes outside of V_1 have zero mass. In particular, this means that even starting from $||x^* - x||_1 = \delta'$ (which would correspond to using $\delta = \delta'/2$ in our analysis), $||x^* - y||_1$ is not bounded in terms of δ', so x^* is not Lyapunov-stable. □

Lemma 4.5 considerably narrows down the set of equilibria for which the question of whether or not they are universally stable remains open.

More specifically, it only remains to consider mass vectors x in which there is a non-trivial connected C of $\mathcal{T}_{\text{act}}(x)$ such that C is a d-regular graph (for some $d \geqslant 1$), is not a locally balanced graph (in particular not a bipartite graph), and for every $u, v \in C$, $x_v = x_u$. We conjecture that such mass vectors are not universally stable; it would follow that in any universally stable equilibrium, all components have size 1, and hence by Proposition 4.4 the active nodes would be at mutual distance 3. Accordingly, we formulate the following:

Conjecture 4.6

Under the local model, a mass vector is a universally stable equilibrium if and only if its active nodes are at pairwise distance at least 3 in the influence graph.

---

[8] Recall that $\mathcal{T}[S]$ denotes the induced subgraph of $\mathcal{T}$ on the node set S.

4.2 Characterization of Lyapunov stable equilibria for $\alpha = 1$

For $\alpha > 1$, we have shown that any equilibrium whose active nodes form an independent set of pairwise node distance at least 3 is Lyapunov-stable. Perhaps surprisingly, this ceases to be true for $\alpha = 1$. Indeed, on the path of length 4, the equilibrium $x^* = (\frac{1}{2}, 0, 0, \frac{1}{2})$ is not Lyapunov-stable.

We can see this instability as follows. Consider vectors of the form $x^{(\delta)} = (\frac{1}{2} - \delta, \delta, \delta, \frac{1}{2} - \delta)$. Under $x^{(\delta)}$, for any $\delta \in (0, \frac{1}{2})$, the interaction mass of nodes 1 and 4 is strictly smaller than the interaction mass of nodes 2 and 3 (whose interaction masses are equal). This implies that no vector $x^{(\delta)}$ can be an equilibrium for $\delta \in (0, \frac{1}{2})$, and that flow will always be directed from nodes 1 and 4 to nodes 2 and 3. Furthermore, the flow from node 1 to node 2 is equal to the flow from node 4 to node 3, implying that at later times, the mass vector will be of the form $x^{(\delta')}$ with $\delta' > \delta$. As we know by Theorem B.1 that the 4-path always converges, the system converges to some mass vector $y = x^{(\delta^*)}$ such that $\delta^* > 0$.[9] Since the update rule is continuous, this y must be an equilibrium. We have proved that the only such equilibrium is the one with $\delta^* = \frac{1}{2}$. Thus, starting from the perturbation $x^{(\delta)}$ of x^*, the system can only converge to a state y in which $y_1 = y_4 = 0$.

While a pairwise distance of 3 between active nodes is not enough to guarantee stability, a pairwise distance of 4 is sufficient.

Theorem 4.7
Let x^* be a mass vector whose active nodes have pairwise distance at least 4. Then, x^* is a Lyapunov-stable equilibrium for $\alpha = 1$.

Proof
The proof is much more involved than the proof of Proposition 2.3, for the following reason: even for arbitrarily small perturbations to x^*, it is possible that inactive neighbors v of an active node u have higher interaction mass; thus, the conditions of Proposition 2.3 do not apply, and in fact, u could lose mass over time. However, we will be able to show that the total mass u loses, starting from a perturbation of magnitude at most δ, is bounded by a function $g(\delta) \to 0$ as $\delta \to 0$.

Let A be the set of all x^*-active nodes, and let $x(0)$ be a perturbation of x^* with $||x^* - x(0)||_1 \leqslant \delta \leqslant \frac{1}{8} \cdot \min_{u \in A} x_u^*$. We will show below that for each node $u \in A$, and all times t, we have that $|x_u(t) - x_u^*| \leqslant 2\delta$. Because

$$\sum_{v \notin A} |x_v(t) - x_v^*| = \sum_{v \notin A} x_v(t) = \sum_{u \in A} (x_u^* - x_u(t)) \leqslant \sum_{u \in A} |x_u(t) - x_u^*|,$$

we obtain that

$$||x(t) - x^*||_1 = \sum_{u \in V} |x_u(t) - x_u^*| \leqslant 2 \sum_{u \in A} |x_u(t) - x_u^*| \leqslant 4n\delta \to 0 \quad \text{as } \delta \to 0.$$

It remains to prove the inequality $|x_u(t) - x_u^*| \leqslant 2\delta$ for all nodes $u \in A$ and times t. Define $W = V \setminus (A \cup \bigcup_{u \in A} T_u)$ to be the set of all nodes at distance at least 2 from all active nodes. We will prove the inequality by showing that any flow from u to its neighbors v can be "charged" against flow from W to v. More formally, we will

---

[9] This also follows directly from our argument with $x^{(\delta)}$.

simultaneously prove the following invariants for all times t and any set $U \subseteq A$:

$$\sum_{w \in W} x_w(t) \leqslant \sum_{w \in W} x_w(0), \tag{7a}$$

$$\sum_{u \in U} x_u(t) \geqslant \sum_{u \in U} x_u(0) - \sum_{w \in W} (x_w(0) - x_w(t)) \quad \text{for any set } U \subseteq A. \tag{7b}$$

Let t be an arbitrary time and assume that the invariants hold at time t. First, we notice some useful consequences of the invariants, including the desired fact that $|x_u(t) - x_u^*| \leqslant 2\delta$.

From Inequality (7a), we get that $\sum_{w \in W} x_w(t) \leqslant \sum_{w \in W} x_w(0) \leqslant \sum_{v \notin A} x_v(0) \leqslant \delta$. Substituting this bound into Inequality (7b) with $U = \{u\}$, and using that $x_u(0) \geqslant x_u^* - \delta$ gives us that $x_u(t) \geqslant x_u^* - 2\delta$. Similarly, using Inequality (7b) with $U = A \setminus \{u\}$ gives us an upper bound of $x_u(t) \leqslant x_u^* + 2\delta$. So we have shown that $|x_u(t) - x_u^*| \leqslant 2\delta$.

Let $(w, v), w \in W, v \notin W$ be an arbitrary edge, and u the unique active neighbor of v in $\mathscr{T}$. The flow on the edge (w, v) is $f_{w \to v}(t) = p \cdot \frac{x_w(t) x_v(t)(\mathscr{M}_v(t) - \mathscr{M}_w(t))}{\mathscr{M}_w(t) \mathscr{M}_v(t)}$. We have just seen that $x_u^* - 2\delta \leqslant x_u(t) \leqslant x_u^* + 2\delta$, so we can also bound $\mathscr{M}_v(t) \geqslant x_u^* - 2\delta$. Applying Inequality (7b) with $U = A \setminus \{u\}$ also gives us an upper bound of $\mathscr{M}_v(t) \leqslant 1 - \sum_{u' \in A, u' \neq u} x_{u'}(t) \leqslant x_u^* + 2\delta$. Furthermore, using the definition of W and Inequality (7b) for $U = A$,

$$\mathscr{M}_w(t) \leqslant \sum_{v \notin A} x_v(t) \leqslant \sum_{v \notin A} x_v(0) + \sum_{w \in W} (x_w(0) - x_w(t)) \leqslant 2\delta.$$

Substituting these bounds, we get that

$$f_{w \to v}(t) \geqslant p \cdot \frac{x_w(t) x_v(t)(x_u^* - 4\delta)}{2\delta(x_u^* + 2\delta)}. \tag{8}$$

By definition of δ, this quantity is always non-negative. In particular, this means that flow goes from w to v; since the edge (w, v) was arbitrary, we have established the invariant (7a).

Next, fix an arbitrary node pair $u \in A, v \in T_u$. The flow from u to v is

$$f_{u \to v}(t) = p \cdot \frac{x_u(t) x_v(t)(\mathscr{M}_v(t) - \mathscr{M}_u(t))}{\mathscr{M}_u(t) \mathscr{M}_v(t)} \leqslant p \cdot \frac{x_u(t) x_v(t) \sum_{w \in W \cap T_v} x_w(t)}{(x_u(t))^2}$$

$$= p \cdot \frac{1}{x_u(t)} \sum_{w \in W \cap T_v} x_w(t) x_v(t) \leqslant p \cdot \frac{1}{x_u^* - 2\delta} \sum_{w \in W \cap T_v} x_w(t) x_v(t).$$

On the other hand, summing the bound (8) overall nodes $w \in W \cap T_v$, we get that

$$\sum_{w \in W \cap T_v} f_{w \to v}(t) \geqslant p \cdot \frac{x_u^* - 4\delta}{2\delta(x_u^* + 2\delta)} \cdot \sum_{w \in W \cap T_v} x_w(t) x_v(t).$$

Because $\delta \leqslant x_u^*/8$, we get that $\frac{1}{x_u^* - 2\delta} \leqslant \frac{x_u^* - 4\delta}{2\delta(x_u^* + 2\delta)}$, so the flow from u to v is at most the total flow from all $w \in W \cap T_v$ to v. For any set $U \subseteq A$, summing this inequality overall $u \in U$ (and noticing that we never double-count the same edge) now shows that the total flow out of U is no more than the total flow out of W; hence, the decrease in U's mass can be charged to a corresponding decrease of mass in W, and we have established Invariant (7b). $\qquad \square$

5 Discussion and conclusions

In this paper, we presented a novel model of cultural dynamics that captures the essential aspect of several previously studied models: the interplay between selection and influence. We concentrated on two instances of this model. In the basic version, the *global model*, each person selects another person from the entire population to interact with. In the *local model*, a person selects an interaction partner from a subset of the population consisting of similar people. We provided a nearly complete treatment of the global model, showing that the system always converges from any initial mass vector, and providing a complete characterization of Lyapunov-stable equilibria.

5.1 Modeling choices

5.1.1 Continuum of individuals

We assumed a continuum of individuals, rather than a finite population. With finite populations, convergence to equilibrium states is quite immediate. The directed Markov Chain of all possible assignments of individuals to nodes of the graph is finite. For any state in which the occupied nodes do *not* form an independent set, there is a sequence of finitely many moves (which has strictly positive probability of occurring) which will result in the individuals being located at an independent set; the latter states are sinks of the Markov Chain. Thus, convergence in finite time is always guaranteed with probability 1. The primary focus of studying such finite Markov Chains would be a focus on the *amount of time* it would take to reach a sink state.

For very large populations, the predictions of the convergence time may be of less interest, and we believe that a focus on the stability of equilibria, and the convergence guarantees in the limit of large populations, are of interest in understanding the outcomes of the influence-selection process.

While several papers such as the early work of Kurtz (1970) and the more refined analysis of Wormald (1999) establish precise connections between discrete-time processes with finite populations and the mean-field continuous-time limit as both time and the population are scaled, we do not believe that these approaches are sufficiently powerful to easily imply results presented here; they may, however, establish that with high probability, the discrete version of the problem stays close to the mean-field approximation. These results also motiviate the continuous-time version of the problem studied here.

5.1.2 Continuous time

In keeping with much of the literature on population dynamics, we treat time as continuous. However, all results proved here hold equally for discrete time; indeed, an earlier version of this paper—still available on the arXiv at http://arxiv.org/abs/1304.7468 (v.1)—carried out all proofs in discrete time. In discrete time, the flows defined in Equation (1) do not correspond to continuous derivatives of the node masses, but rather to discrete changes from step t to $t + 1$. The proofs of all results stay essentially the same, requiring only the obvious modifications. The main

exceptions are Lemma 3.3 and Theorem 3.4, whose proofs become slightly more intricate: a simple continuity argument cannot be applied, as discrete changes may lead to jumps in the $x_u(t)$. However, careful accounting and judicious choices of interval sizes δ still make the proofs go through.

One argument frequently raised against discrete-time analysis is that it may lead to oscillations in states, in particular when states are updated synchronously. In our proofs, we have not observed such oscillations, and indeed conjecture that both the discrete-time and continuous-time versions will always converge to an equilibrium. In the cases where a proof of this convergence has been difficult, this difficulty has persisted in both continuous and discrete time.

5.1.3 Discrete graph structures

Throughout, we have assumed that the interaction and influence graphs are "discrete," in the sense that the propensity of individuals to interact with (or be influenced by) individuals of adjacent types is the same for all adjacent types. In reality, the world will not be as black-and-white; rather, there will be some types v adjacent to u that are more likely than others to succeed in convincing individuals from node u to switch. For instance, in the introductory example from Section 1, while radical protestants may be most likely to become moderate protestants (or stay radical), a small fraction may directly become atheists.

In a more general form of the model, the probability p for switching between types would depend on the specific types, i.e. be of the form $p_{u,v}$, where it is expressly possible that $p_{u,v} \neq p_{v,u}$. $p_{u,v} = 0$ would correspond to the absence of an edge from u to v. Similarly, we could assign weights to the edges of the interaction graph, and have meeting probabilities follow those weights.

This more general model becomes significantly more complex to analyze. When $p_{u,v} = p_{v,u}$ for all pairs (u,v), much of the analysis in the present work carries over, but for asymmetric versions, a more complex approach may be needed. Similarly, given the difficulties caused even by the simple local model, a more general weighted interaction graph model looks like a rather formidable challenge.

5.2 Open questions

An open question is to predict the equilibrium to which the system converges starting from a given initial mass vector. We suspect that with probability 1 over possible starting states, the system converges to an equilibrium in which the active nodes form an independent set.

The local model involves, at its heart, a dynamical system on the population fractions that is complicated even for small numbers of variables. As such, it raises many interesting and challenging questions, and we have made progress on some of these. In particular, we know that on paths of length 3 (for $\alpha > 1$) and at most 5 (for $\alpha = 1$), the system converges from any starting state. However, it is open whether convergence occurs for all graphs. On the stability frontier, for $\alpha > 1$, we conjecture that the only Lyapunov-stable equilibria are those in which the active nodes have pairwise distance at least 3. We showed that such equilibria are indeed Lyapunov-stable, and that a number of other equilibria are not Lyapunov-stable—including

ones in which the active nodes form any other independent set, or ones in which they form a *locally balanced graph* (a class that includes bipartite graphs). Finally, we would like to raise an even more challenging question: does the dynamical system defined by the general model—or even the general model—always converge?

Acknowledgments

References

Abrams, D. M., & Strogatz, Steven H. (2003). Modelling the dynamics of language death. *Nature*, **424**, 900.

Anagnostopoulos, A., Kumar, R., & Mahdian, M. (2008). Influence and correlation in social networks. In *Proceedings of the 14th ACM sigkdd International Conference on Knowledge Discovery and Data Mining*, pp. 7–15.

Aral, S., Muchnik, L., & Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences USA*, **106**(51), 21544–21549.

Axelrod, R. (1997). The dissemination of culture. *Journal of Conflict Resolution*, **41**(2), 203–226.

Bakshy, E., Rosenn, I., Marlow, C. A., & Adamic, L. A. (2012). The role of social networks in information diffusion. In *Proceedings of the 21st International World Wide Web Conference*.

Ben-Naim, E., Krapivsky, P., & Redner, S. (2003). Bifurcations and patterns in compromise processes. *Physica D*, **183**(3), 190–204.

Bramoullé, Y., Djebbari, H., & Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, **159**, 41–55.

Brandes, U., & Erlebach, T. (eds). (2005). *Network analysis: Methodological foundations*. Berlin/Heidelberg: Springer-Verlag.

Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics*, **81**, 591–646.

Centola, D., Gonzalez-Avella, J. C., Eguiluz, V. M., & San Miguel, M. (2007). Homophily, cultural drift, and the co-evolution of cultural groups. *Journal of Conflict Resolution*, **51**(6), 905–929.

Cohen, J. M. (1977). Sources of peer group homogeneity. *Sociology in Education*, **50**(Oct.), 227–241.

Deffuant, G., Neau, D., Amblard, F., & Weisbuch, G. (2000). Mixing beliefs among interacting agents. *Advances in Complex Systems*, **3**, 87–98.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, **486**, 75–174. Eprint arXiv: 0906.0612.

Handcock, M. S., Raftery, A. E., & Tantrum, J. M. (2007). Model-based clustering for social networks (with discussion). *Journal of the Royal Statistical Society, Series A*, **170**, 301–354.

Hegselmann, R., & Krause, U. (2002). Opinion dynamics and bounded confidence: Models, analysis and simulation. *Journal of Artificial Societies and Social Simulation*, **5**(3), 1–24.

Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, **97**, 1090–1098.

Kandel, D. B. (1978). Homophily, selection, and socialization in adolescent friendships. *American Journal of Sociology*, **84**(2), 427–436.

Kurtz, T. G. (1970). Solutions of ordinary differential equations as limits of pure jump markov processes. *Journal of Applied Probability*, **7**(1), 49–58.

LaFond, T., & Neville, J. (2010). Randomization tests for distinguishing social influence and homophily effects. In*Proceedings of the 19th International World Wide Web Conference*, pp. 601–610.

McPherson, M., Smith-Lovin, L., & Cook, J M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, **27**, 415–444.

Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. New York, NY: Penguin Press.

Patriarca, M., & Leppanen, T. (2004). Modeling language competition. *Physica A*, **338**(1–2), 296–299.

Poole, K., & Rosenthal, H. (1991). Patterns of congressional voting. *American Journal of Political Science*, **35**(1), 228–278.

Schaeffer, S. E. (2007). Graph clustering. *Computer Science Review*, **1**(1), 27–64.

Shalizi, C. R., & Thomas, A. C. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods and Research*, **40**, 211–239.

Snijders, T. A. B., Steglich, C., & Schweinberger, M. (2007). Modeling the co-evolution of networks and behavior. In K. van Montfort, H. Oud, & A. Satorra (Eds.), *Longitudinal models in the behavioral and related sciences* (pp. 41–71). Mahwah, NJ: Lawrence Erlbaum.

Spector, D. (2000). Rational debate and one-dimensional conflict. *Quarterly Journal of Economics*, **115**(1), 181–200.

Stauffer, D., Castello, X., Eguiluz, V. M., & Miguel, M. S. (2007). Microscopic Abrams-Strogatz model of language competition. *Physica A*, **374**(2), 835–842.

Steglich, C., A.B.Snijders, Tom, & Pearson, M. (2010). Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology*, **40**, 329–393.

Sunstein, C. R. (2009). *Republic.com 2.0*. Princeton, NJ: Princeton University Press.

Wormald, N. C. (1999). The differential equation method for random graph processes and greedy algorithms. In M. Karonski, & H. Proemel (Eds.), *Lectures on approximation and randomized algorithms* (pp. 73–155). PWN.

Appendix A: Convergence on a 3-node path (Proof of Theorem 4.1)

Observe that the flow between two types u and v does not necessarily go in the same direction at all times, but instead may change its direction. To keep track of the changes in direction, we define a *configuration* of the system to be a labeling of all edges (u, v) in $\mathscr{T}$ by the direction along which flow is traveling (i.e. whether it travels from u to v or from v to u). A configuration which cannot transition to any other configuration is called a *sink configuration*. Sink configurations are important because they guarantee convergence by Lemma 2.2.

In the case of a 3-node path and $\alpha \geqslant 2$, there are four possible configurations. We study transitions among the configurations as the system evolves over time; we show that each configuration is either a sink configuration, or it has the property that any change in the direction of an edge leads to a sink configuration. This ensures that there can be at most one change in the direction of flow as the system evolves.

For a 3-node path and $\alpha < 2$, convergence will follow as a special case of the more general Theorem A.2, which establishes convergence for all star graphs when $\alpha < 2$. For a 3-node path and $\alpha \geqslant 2$, we prove the following lemma (which, jointly with Lemma 2.2, implies Theorem 4.1 for $\alpha \geqslant 2$).

Lemma A.1
Consider the local model with $\alpha \geqslant 2$ such that the influence graph $\mathscr{T}$ is a 3-node path. Then, there is a time t_0 such that for any $t > t_0$, no flow changes its direction.

Proof
Let the nodes of the path be $(1, 2, 3)$, in order. Consider an arbitrary time t, and recall that $\mathscr{M}_1(t) = \alpha x_1(t) + x_2(t)$. Node 1's interaction mass is decreased at a rate of $\alpha f_{1 \to 2}(t)$ from flow leaving node 1 to node 2, and increased at a rate of $f_{1 \to 2}(t) + f_{3 \to 2}(t)$ from flow entering node 2. By applying the same reasoning to nodes 2 and 3, we get

$$
\begin{aligned}
\dot{\mathscr{M}}_1(t) &= f_{3 \to 2}(t) - (\alpha - 1)f_{1 \to 2}(t), \\
\dot{\mathscr{M}}_2(t) &= (\alpha - 1)(f_{1 \to 2}(t) + f_{3 \to 2}(t)), \\
\dot{\mathscr{M}}_3(t) &= f_{1 \to 2}(t) - (\alpha - 1)f_{3 \to 2}(t).
\end{aligned}
\tag{A 1}
$$

Let $x_i = x_i(t), \mathscr{M}_i = \mathscr{M}_i(t), f_{i \to j} = f_{i \to j}(t)$ for $i, j = 1, 2, 3$. We will distinguish three cases based on the relative sizes of $\mathscr{M}_1, \mathscr{M}_2, \mathscr{M}_3$.

1. If $\mathscr{M}_2 \geqslant \mathscr{M}_1$ and $\mathscr{M}_2 \geqslant \mathscr{M}_3$, then both $f_{1 \to 2}$ and $f_{3 \to 2}$ are non-negative. According to Equation (A 1), $\mathscr{M}_2$ increases by at least as much as both $\mathscr{M}_1$ and $\mathscr{M}_3$, so the same inequality will subsequently as well. Thus, we have reached a sink configuration.
2. If $\mathscr{M}_2 < \mathscr{M}_1$ and $\mathscr{M}_2 < \mathscr{M}_3$, then both $f_{1 \to 2}$ and $f_{3 \to 2}$ are negative. By Equation (A 1), $\mathscr{M}_2$ decreases by at least as much as both $\mathscr{M}_1$ and $\mathscr{M}_3$, so again, the inequalities will hold forever, and we have reached a sink configuration.
3. The remaining case is that $\mathscr{M}_2 < \mathscr{M}_1$ and $\mathscr{M}_2 \geqslant \mathscr{M}_3$. (The case $\mathscr{M}_2 < \mathscr{M}_3, \mathscr{M}_2 \geqslant \mathscr{M}_1$ is symmetric.) Here, $\mathscr{M}_3$ decreases, $\mathscr{M}_1$ increases, and $\mathscr{M}_2$ may increase or decrease. If the relative order of $\mathscr{M}_1, \mathscr{M}_2, \mathscr{M}_3$ stays the same for all times after t, then we have reached a sink configuration. Otherwise, at some time $t' \geqslant t$, we must reach either a configuration with $\mathscr{M}_2(t') < \mathscr{M}_1(t'), \mathscr{M}_2(t') < \mathscr{M}_3(t')$ or with $\mathscr{M}_2(t') \geqslant \mathscr{M}_1(t'), \mathscr{M}_2(t') \geqslant \mathscr{M}_3(t')$. Either of those configurations is a sink configuration by the preceding two cases.

In summary, each configuration is either a sink configuration, or will reach a sink configuration at the next transition to a different order of interaction masses.   □

Next, we prove that for $\alpha < 2$ the process converges on every star graph (and in particular on the 3-path).

Theorem A.2
Under the local model with $\alpha < 2$, if the influence graph is a star graph, then the system converges from any starting state.

In the remainder of this section, we prove Theorem A.2. More specifically, we show that eventually the system enters a sink configuration. The first lemma towards the proof holds for arbitrary values of α.

Lemma A.3
Consider the local model with an arbitrary $\alpha \geqslant 1$ such that the influence graph $\mathscr{T}$ is a star graph. Then, at any time, the number of edges with flow directed away from the center is at most $\lfloor \alpha \rfloor$.

Proof

Denote the central node by u. Fix some time t, and let R be the set of all peripheral nodes v such that the flow on the edge (u, v) is directed from u to v. Because the flow is directed towards v, $\mathscr{M}_v > \mathscr{M}_u$ for all $v \in R$. Rearranging this inequality gives us that $(\alpha - 1)(x_v - x_u) > \sum_{w \neq u,v} x_w$, which implies in particular that $x_v > \frac{\sum_{w \neq u,v} x_w}{\alpha - 1}$. Summing overall $v \in R$ now implies that

$$\sum_{v \in R} x_v > \sum_{v \in R} \frac{\sum_{w \neq u,v} x_w}{\alpha - 1} \geqslant (|R| - 1) \frac{\sum_{v \in R} x_v}{\alpha - 1}.$$

Thus, we have that $|R| \leqslant \lfloor \alpha \rfloor$. □

Lemma A.4

Consider the local model such that the influence graph $\mathscr{T}$ is a star graph, and with $\alpha < 2$. Then, any configuration in which flow on exactly one edge is directed away from the center node is a sink configuration.

Proof

Let u be the center node. Suppose that at time t, the system is in a configuration in which the flow on exactly one edge (u, v) is directed away from the center; so $\mathscr{M}_u(t) < \mathscr{M}_v(t)$. By Lemma A.3, there can be at most one edge on which the flow is directed away from u, and (u, v) is such an edge. The changes in interaction masses are

$$\dot{\mathscr{M}}_u(t) = (\alpha - 1) \sum_{w \neq u,v} f_{w \to u}(t) - (\alpha - 1) f_{u \to v}(t)$$

$$\dot{\mathscr{M}}_v(t) = (\alpha - 1) f_{u \to v}(t) + \sum_{w \neq u,v} f_{w \to u}(t).$$

Their difference is

$$\dot{\mathscr{M}}_u(t) - \dot{\mathscr{M}}_v(t) \leqslant (\alpha - 2) \sum_{w \neq u,v} f_{w \to u}(t) - 2(\alpha - 1) f_{u \to v}(t).$$

Because $\alpha < 2$, the right-hand side is negative, so u's interaction mass grows more slowly (or decreases faster) than v's, implying that the edge (u, v) remains directed from u to v. Hence, the configuration is a sink configuration. □

Theorem A.2 now follows from Lemmas 2.2, A.3, and A.4, as follows. If the system ever enters a configuration in which exactly one edge has flow directed away from the center, then by Lemma A.4, it subsequently stays in this configuration forever, so by Lemma 2.2, the system converges. By Lemma A.3, the only other alternative is that the system is always in the configuration with all edges directed inwards; then, again, it converges by Lemma 2.2.

Appendix B: Local model with $\alpha = 1$: Convergence on a path

Assume that the active subgraph is an n-node path with nodes $(1, 2, \dots, n)$. The endpoints of the path, nodes 1 and n, always have interaction masses no larger than their neighbors (nodes $2, n - 1$), implying that their masses $x_1(t), x_n(t)$ are monotonically non-increasing. This implies convergence of $x_1(t), x_n(t)$ as $t \to \infty$. In the following proposition, we will exploit the convergence at the endpoints to show

that $x_2(t)$ and $x_{n-1}(t)$ must also converge. For a path of length at most 5, this implies convergence of the vector x to an equilibrium, as the total mass stays constant. Our technique does not apply beyond length 5; we do not know of a direct way to generalize the argument inductively to paths of arbitrary lengths.

Theorem B.1

Consider the local model with $\alpha = 1$. If the influence graph is a path of $n \leqslant 5$ nodes, then the system converges.

Proof

We already argued above that $x_1(t)$ and $x_n(t)$ converge. If the path has three nodes, then $x_2(t) = 1 - x_1(t) - x_3(t)$ (by mass conservation), so $x_2(t)$ converges as well. So assume that $n \in \{4, 5\}$. Below, we show that $x_2(t)$ converges as well; a symmetric argument applies to $x_{n-1}(t)$. If the path has four nodes, we are done at this point. If the path has five nodes, then $x_3(t) = 1 - x_1(t) - x_2(t) - x_4(t) - x_5(t)$ must converge as well. Thus, $x(t)$ converges in all cases.

To prove that $x_2(t)$ converges, we distinguish two cases, based on $y_1 = \lim_{t \to \infty} x_1(t)$.

1. If $y_1 = 0$, there are two subcases. If $x_1(t) \geqslant x_2(t)$ for all t, then clearly, $x_2(t) \to 0$ as well. Otherwise, there exists a t_0 with $x_2(t_0) > x_1(t_0)$. By Equation (2), specialized to the local model and $\alpha = 1$, we obtain that for any t,

$$\dot{x}_1(t) = p \cdot x_1(t) \cdot \left(\frac{x_1(t)}{\mathcal{M}_1(t)} + \frac{x_2(t)}{\mathcal{M}_2(t)} - 1 \right),$$

$$\dot{x}_2(t) = p \cdot x_2(t) \cdot \left(\frac{x_1(t)}{\mathcal{M}_1(t)} + \frac{x_2(t)}{\mathcal{M}_2(t)} + \frac{x_3(t)}{\mathcal{M}_3(t)} - 1 \right).$$

Then, clearly, $x_1(t) < x_2(t)$ implies $\dot{x}_1(t) < \dot{x}_2(t)$. In particular, this means that $x_2(t) > x_1(t)$ for all $t \geqslant t_0$. In turn, this inequality is used in the last step of the following derivation:

$$\max(f_{3 \to 2}(t), 0) \leqslant p \cdot \frac{x_2(t) \, x_3(t)}{\mathcal{M}_2(t) \, \mathcal{M}_3(t)} \cdot x_1(t)$$

$$= p \cdot \frac{x_1(t) \, x_2(t) \, x_3(t)}{\mathcal{M}_1(t) \, \mathcal{M}_2(t)} \cdot \frac{\mathcal{M}_1(t)}{\mathcal{M}_3(t)}$$

$$= f_{1 \to 2}(t) \cdot \frac{x_1(t) + x_2(t)}{x_2(t) + x_3(t) + x_4(t)}$$

$$\leqslant 2 f_{1 \to 2}(t).$$

Thus, the total amount of flow entering node 2 after time t_0 is at most $3 \int_{t_0}^{\infty} f_{1 \to 2}(t) dt \leqslant 3 x_1(t_0)$. The reason for the last inequality is that flow never enters node 1, so the total amount of flow that can leave node 1 for node 2 after t_0 is at most the amount that was at node 1 at time t_0.

Let $F^+(t)$ (resp., $F^-(t)$) be the total amount of flow that has entered (resp., left) node 2 up to time t. We have just proved that $F^+(t) - F^+(t_0) \leqslant 3x_1(t_0)$. Therefore, $F^+(t)$, being monotone and bounded, must converge. Because flow can only leave node 2 when it was already there, we get that $F^-(t) \leqslant x_2(0) + F^+(t)$ is also bounded, and must also converge. Hence, $x_2(t) = x_2(0) + F^+(t) - F^-(t)$, being the difference between two convergent functions, must also converge.

2. If $y_1 > 0$, we will pursue a similar argument, but this time bounding the cumulative flow *out of* node 2 instead of into it. Because $x_2(t) \leq 1$ for all times t, this means that the cumulative flow into node 2 must also be bounded. Then, an identical argument to the previous paragraph shows that $x_2(t) = x_2(0) + F^+(t) - F^-(t)$ must converge.

No flow can ever leave node 2 for node 1, so we just need to bound the flow from node 2 to node 3. Flow leaves node 2 for node 3 at time t if and only if $x_4(t) \geq x_1(t)$. Since $x_1(t) \geq y_1$, it follows that $\mathscr{M}_2(t), \mathscr{M}_3(t) \geq y_1$ as well. Therefore, we can bound the non-negative flow from node 2 to node 3 as follows:

$$
\begin{aligned}
\max(f_{2 \to 3}(t), 0) &\leq p \cdot \frac{x_2(t) \, x_3(t)}{\mathscr{M}_2(t) \, \mathscr{M}_3(t)} \cdot x_4(t) \ \leq \ p \cdot \frac{x_2(t) \, x_3(t)}{y_1^2} \\
&\leq p \cdot \frac{x_1(t) \, x_2(t) \, x_3(t)}{y_1^3} \ \leq \ \frac{1}{y_1^3} \cdot p \cdot \frac{x_1(t) \, x_2(t) \, x_3(t)}{\mathscr{M}_1(t) \mathscr{M}_2(t)} \ = \ \frac{1}{y_1^3} \cdot f_{1 \to 2}(t).
\end{aligned}
$$

Thus, the total positive flow from node 2 to node 3 is bounded above by a constant times the total flow from node 1 to node 2, which in turn is at most $x_1(0)$. $\qquad \square$