Engineering Agreement: The Naming Game with Asymmetric and Heterogeneous Agents

Jie Gao[1] and **Bo Li**[2] and **Grant Schoenebeck**[2] and **Fang-Yi Yu**[2] *
[1]Stony Brook University; [2]University of Michigan

Abstract

Being popular in language evolution, cognitive science, and culture dynamics, the Naming Game has been widely used to analyze how agents reach global consensus via communications in multi-agent systems. Most prior work considered networks that are symmetric and homogeneous (e.g., vertex transitive). In this paper we consider asymmetric or heterogeneous settings that complement the current literature: 1) we show that increasing asymmetry in network topology can improve convergence rates. The star graph empirically converges faster than all previously studied graphs; 2) we consider graph topologies that are particularly challenging for naming game such as disjoint cliques or multi-level trees and ask how much extra homogeneity (random edges) is required to allow convergence or fast convergence. We provided theoretical analysis which was confirmed by simulations; 3) we analyze how consensus can be manipulated when stubborn nodes are introduced at different points of the process. Early introduction of stubborn nodes can easily influence the outcome in certain family of networks while late introduction of stubborn nodes has much less power.

1 Introduction

The analysis of shared conventions in multi-agent systems and complex decentralized social networks has been the focus of study in several diverse fields, such as linguistics, sociology, cognitive science, and computer science. The problem of how such conventions can be established, from among countless options, without a central coordinator has been addressed by several disciplines (Nowak and Krakauer 1999; Brighton and Kirby 2001). Among them, the multi-agent models and mathematical approaches gain the most attention by accounting for both the network topology and opinion change over time (Steels 2005; Nowak, Plotkin, and Jansen 2000; Baronchelli, Loreto, and Steels 2008; Pickering and Lim 2016; Franks, Griffiths, and Jhumka 2013). It has been shown that the emergence of new political, social, economic behaviors, and culture transmission are highly dependent on such convention dynamics (Backstrom et al. 2006; Hurford 1989; Nowak, Plotkin, and Krakauer 1999).

In order to analyze the social dynamics in multi-agent systems in depth, we focus on one stylized model, the Naming Game, in which agents negotiate conventions through local pairwise interactions (Steels 1995; Baronchelli et al. 2006a). The Naming Game captures the generic and essential features of an agreement process in networked agent-based systems. Briefly speaking, when two agents wish to communicate, one agent, the speaker, randomly selects one convention from her list of current conventions and uses this convention to initiate communication with the listener. If the listener recognizes that convention, both the speaker and listener purge their lists of current conventions to only include that "successful" convention. If the listener does not recognize that convention, she adds it to her list of known conventions.

This simple model is able to account for the emergence of shared conventions in a homogeneous population of agents. Both simulations and experiments have been conducted on various network topologies. However many key questions, especially those related to asymmetric and heterogeneous agents, remain open. For example: what network topologies enable the fastest convergence? Does community structure help or harm convergence? Does homogeneity or heterogeneity help or harm convergence? How robust are the dynamics to possible manipulations by a small number of agents? Moreover, rigorous theoretical analysis is almost entirely absent in previous work on the Naming Game. In this paper we aim to fill in the literature in the following aspects:

1. We discovered that the star graph empirically converges faster than all previously considered graphs for the Naming Game. This network differs from previously analyzed topologies in that it is not symmetric (vertex transitive). In some sense, it is not too surprising that the star graph, an asymmetric graph, works so well to reach consensus, which is a symmetry breaking problem. Though, from first principles, this is far from obvious, and other asymmetric graphs, for example a multi-level tree, perform extremely poorly.

2. To understand network topologies that inhibit fast convergence of the Naming Game, we study two networks with community structures: agents divided into two dis-

*J. Gao would like to acknowledge support through NSF DMS-1418255, CCF-1535900, CNS-1618391 and AFOSR FA9550-14-1-0193. The remaining authors are pleased to acknowledge the support of the NSF via AitF-1535912 and CAREER-1452915.

connected communities; and a multi-level tree. For the first network, it is clear that it cannot converge to consensus (it is disconnected). We investigate how much inter community communication needs to be added in order to facilitate convergence. Empirically we observe a sharp threshold on the level of inter community communication: above this threshold, fast convergence is guaranteed, and below it the dynamics fail to converge before time out. We give theoretical justifications for this threshold by showing that convergence takes exponentially long if inter community communication is insufficient (below the threshold). For the second network, the multi-level tree, we observe via simulations that it converges exceedingly slowly—we conjecture that it takes exponential time. For this network, we perform the same simulation tests for adding homogeneity and obtain similar results.

We show that with added communication, the community divisions that thwart consensus can be overcome. Perhaps surprisingly, the amount of intercommunity communication required after disagreement is entrenched, is not substantially more than the amount of communication needed to avoid such division in the first place.

3. Finally, we analyze a third way of introducing asymmetry and heterogeneity: including "stubborn" nodes that do not follow the standard Naming Game protocol. Our experimental results suggest the following hypothesis: in some graphs (e.g. cliques) even a small constant (e.g. 5) number of stubborn nodes can assure convergence to a particular name. However, in others networks (e.g. star graphs, grid graphs, Kleinberg's small world models), the number of nodes required seems to grow with the size of the graph. Additionally, we prove that in a complete graph, manipulation after convergence is much harder than before: there exists a value p such that if an adversary controls more than a p fraction of the nodes, consensus results can be easily manipulated; otherwise it takes exponential time to manipulate the consensus.

   The results on stubborn nodes have implications for the use of the Naming Game in distributed systems. In Steels and McIntyre (1998) it was assumed that the protocol would be robust to manipulation. We confirmed this claim if the stubborn nodes appear after the system has converged. But in certain networks these protocols are immensely vulnerable to rogue agents that appear from the start.

Related Work

Baronchelli et al. (2006b) proposed the Naming Game as a simple multi-agent framework that accounts for the emergence of shared conventions in a structured population. One of the most important problems for Naming Game is to understand how fast the global consensus can be reached and what factors affect it. Some research has been conducted to analyze the effect of network topology on the Naming Game dynamics (Dall'Asta et al. 2006). Lu, Korniss, and Szymanski (2009) show via simulations on real-world graphs that communities show speedy convergence of the dynamics. Centola and Baronchelli (2015), using human-subject study, empirically demonstrate the spontaneous creation of universally adopted social conventions and show simple changes in a population's network structure can greatly change the dynamics of norm formation. Baronchelli et al. (2007) show that finite connectivity, combined with the small-world property, ensures superior performance in terms of memory usage and convergence rate to that of the grid or complete network. Additionally, a dynamically evolving topology of co-evolution of language and social structure has been studied by Gong et al. (2004), for a more complex language game.

One common way to influence the social dynamics and facilitate the converging process towards the consensus is to break the symmetry. Lu, Korniss, and Szymanski (2009)Lu et al. have made use of a special kind of agents called "committed" nodes, who will stick to a preferred opinion without deviating, and show that such agents often reduce the time needed to reach consensus. However, in their work they did not evaluate how these nodes might influence which name was converged upon. Additionally, they did not study how the network topology interacted with stubborn nodes or how robust the communication protocol is.

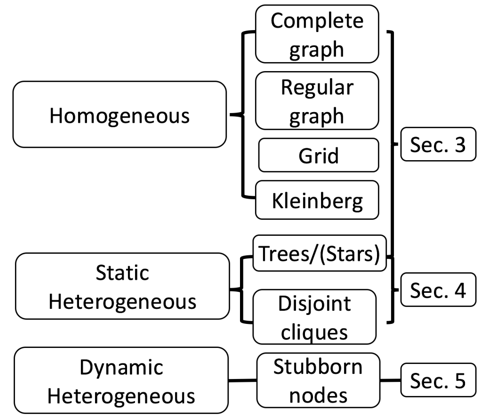


Figure 1: Overview of considered graph structures.

2 Preliminaries

We present here the version of the *Naming Game* introduced in Baronchelli et al. (2006a) in which agents negotiate conventions (names), i.e. associations between forms and meaning. The process stops when all agents reach consensus on a single 'name.' The Naming Game is played by agents on a (weighted) graph $G = (V, E, w)$ and proceeds in steps. At each step t, each agent v, is characterized by its inventory (list of names) $A_t(v) \subseteq \mathbb{S}$. At time 0 each agent has an initial inventory $A_0(\cdot)$ which is possibly empty. At each time step $s = 1, 2...$

- An edge is randomly chosen with probability proportional to its weight; and with equal chance one vertex incident to the edge is considered as the speaker and the other as the listener.
- The speaker v selects a word c uniformly at random from its inventory $A_t(v)$ and sends c to the listener u. If the

speaker's inventory is empty, the speaker invents a new word c (one that is not in the list of any other agent).

- If the word is in the listener's inventory, $c \in A_s(u)$, the interaction is a "success", and both the speaker and listener remove all words besides c from their inventories.
- If the word is not in the listener's inventory, $c \notin A_s(u)$, the interaction is a "failure" and the listener adds c to its inventory.

The process stops when all the inventories are a singleton of the same name, and we say the process has reached consensus. Notice that the only time a node can have an empty inventory is if it starts that way and has yet to engage in any interaction.

The way in which agents may interact with each other is determined by the topology of the underlying contact network. Here we will introduce the models considered in this paper.

1. *Complete graphs*: all agents are mutual connects.
2. *Regular random graph* $G_{n,k}$ (see Bollobás (1998)): every node has degree $k = 8$ and the connection is randomly sample under this constrain.
3. *Kleinberg's small world model* (Kleinberg 2000): in standard Kleinberg's model the nodes are on two dimensional grid. Each node u connects to every other node within Manhattan distance p as strong ties, and there are q weak ties which connects to other nodes v proportional to $d(u, v)^\alpha$. In our simulation, the each nodes has 4 strong tie which is $p = 1$, and 4 weak ties with $\alpha = 2$.
4. *Watts-Strogatz's small world model* (Watts and Strogatz 1998): the nodes are on one-dimensional ring, and connect to 8 nearest nodes with respect to Manhattan distance, then we rewire the edges of independently with probability 0.5.
5. *Complete bipartite graph* is a bipartite graph such that every pair of graph vertices in the two sets are adjacent. If there are p and q graph vertices in the two sets, the complete bipartite graph is denoted $K_(p, q)$.
6. The *trees* in this paper refer to perfect k-ary trees with height h—that is, a rooted tree with h levels where each node except leaf nodes has exactly k children and the leaf nodes are all at the level h. Note that a *star graph* with n leaves is the complete bipartite graph $K_{1,n}$. Alternatively, a star graph can also be defined as rooted tree of branching factor $n - 1$ with depth 1.

3 Networks with Fast and Slow Convergence

In this section we study the convergence rate of various graphs. Here we show that a family of asymmetric graphs, the star graphs, empirically converge faster than previously proposed graphs. Next, we point out, perhaps surprisingly, that the convergence time of a multi-level tree is extremely slow. We will engineer and analyze fast converge versions of trees by adding random edges in Section 4.

We first examine the convergence time for different graphs on a large scale. Here we calculate the time in terms of the number of communication steps denoted as "s". We look at complete graphs, random regular graphs ($G_{n,k}$ graphs), Kleinberg's small world graphs, Watts-Strogatz graphs, as well as star and tree graphs. Unless mentioned otherwise, we will use the same setting defined above in Section 2. From Figure 2 we can see that the star graph converges the fastest. The tree graph is in fact the slowest. If the tree has two levels with 5000 nodes, after 10^7 steps the nodes still cannot reach consensus. Therefore we did not present the consensus time of the tree in the figure. Among the rest of the graphs, the Kleinberg's small world model is the second slowest, while the other graphs have convergence rate roughly a constant factor of each other.

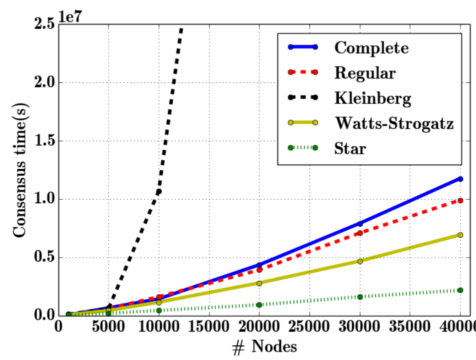


Figure 2: Evaluation of the consensus time for different graphs with size growing until 40000.

The network topology's impact on the Naming Game's consensus time is fairly intriguing. To better understand the results, let us consider the best and worst topology scenarios for multiple agents to reach consensus. The best (quickest) way to reach consensus is to have a specific node to inform all the other nodes of the name. In other words, it is represented by a star graph and the center node is always the speaker. In the naming game framework, even when the speaker/listener role assignment is uniformly random, the star graph is still the fastest in reaching global consensus. This is partly attributed by the asymmetry inherent in the star graph topology.

To analyze the effect of asymmetry, we simulate the graph morphing from a balanced complete bipartite graph to a star by increasing the number of vertices in the larger side of a complete bipartite graph. Figure 3 shows the converge time for various complete bipartite graphs. Moving to the right in the figure, the graph becomes more asymmetric and we see that the convergence time decreases. Note that at $m = n$ ($m/n = 1$), this is a balanced bipartite network, and at $m = 2n - 1$ ($m/n \approx 2$) this is a star graph. This finding is also aligned with the idea that breaking symmetry can improve consensus efficiency for naming game via "stubborn" agents (Lu, Korniss, and Szymanski 2009) (and see Section 5).

On the other hand, the worst graph topology for reaching global consensus is the multi-level tree graph. We hypothesize that this is due to the "community structure" embedded

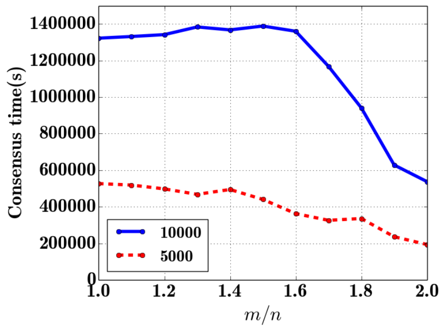


Figure 3: Evaluation for converging time for various complete bipartite graphs $K_{m,2n-m}$ where m is the cardinality of the larger partition of vertices.

in the tree that converge fast by themselves. In a two-level tree, the subtrees of the main tree are themselves star graphs. Such community structure enables fast "local" convergence of the dynamics within the communities, but face challenges in reaching global convergence — the communities are trying to influence each other but each community has more internal influence than external influence. This phenomena is the topic of the next section, where we give both empirical and rigorous theoretical analysis.

4 Effects of Community Structure

In this section we study the effects of community structure using two network models, one of them is a dense graph and the other one a sparse graph. The first is a graph of heterogeneous agents divided into two disconnected communities. The simplicity of this model permits theoretical analysis of precisely how and when community structure can exhibit convergence. The second is a multi-level tree introduced in the previous section.

Given a weighted graph G where the sum of the weights is W we construct $\mathrm{Hom}(G,p)$ by adding $\frac{p}{1-p}\frac{W}{\binom{n}{2}}$ mass to each edge (creating a new edge if it does not exist). This effectively samples the complete graph with probability p and the graph G with probability $1-p$.

For each network, we first examine the convergence rate of $\mathrm{Hom}(\cdot,p)$ using simulations. We show that adding a sufficient amount of homogeneity overcomes the heterogeneity. For the first network, we will provide a theoretical analysis which predicts, supports, and explains the empirical results.

Disjoint Cliques

Naturally, a graph G of $2n$ heterogeneous agents divided into two equally sized disconnected communities will not converge to consensus. As p increases from 0 toward 1 $\mathrm{Hom}(G,p)$ becomes a network of increasingly interconnected communities.

Additionally, the behavior of the Naming Game depends on the initial states, i.e., the collection of names at these nodes at the beginning. We consider two situations for the initial states. 1) "Empty" start, where all nodes start with empty lists $\forall v \in V, A_0(v) = \phi$. 2) "Segregated" start, in which the two groups have different initial opinions, $\forall v \in V_1, A_0(v) = \{0\}$ and $\forall v \in V_2, A_0(v) = \{1\}$. Clearly it is more challenging for the Naming Game to reach global convergence under the segregated initial state.

Simulation Results. Figure 4 (row 1 (a)) shows the convergence time for different values of p under different initial scenarios on graphs of size n. For each setting we run the simulation multiple times and plot the time to reach consensus for each run as a dot in the figure. In certain situations it is hard to reach consensus even after a long time. Therefore we set 10^7 as the time-out criteria – i.e., if no consensus is reached after 10^7 rounds and we stop the simulation. From Figure 4 (row 1 (a)) we can see that when p is smaller it is harder to reach consensus for all situations. When p is sufficiently small all situations may hit the timeout condition before consensus is reached. In addition, the threshold of p which allows this happen is larger for the "segregated" initial setup compared to the empty initial setup. Similarly, for graphs of larger size it is easier to hit the time out condition. When $p > 0.2$ the time to reach consensus for all situations is small so we chose not to plot it.

To further analyze the naming game behavior when p is in between $[0, 0.25]$, we show in Figure 4 (row 2 (a)) the fraction of trials failing to reach consensus (before timing out) with different values of p. It is clear that for the empty start initial condition, the game will time out at about $p = 0.24$, while for the segregated start case, the game will time out when p is around 0.26. This threshold value changes with the size of the local community.

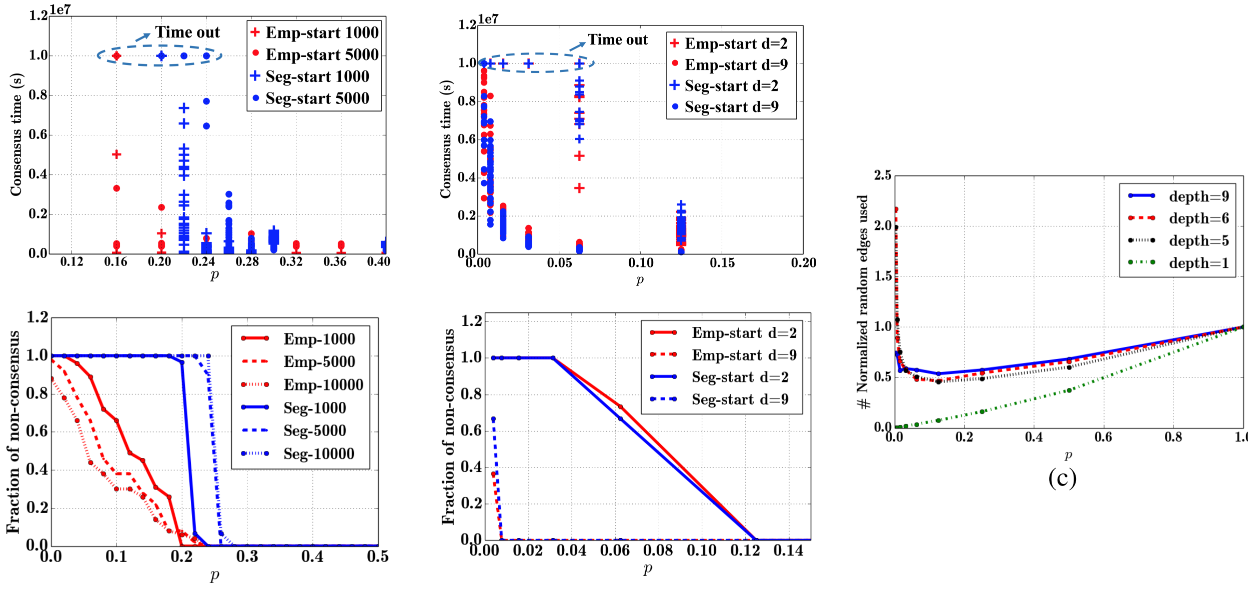
Curiously, for the "empty" start, graphs with smaller sizes are more likely to encounter timeouts than their larger counterparts. This may be because the smaller size of each community results in a greater chance of quickly reaching local consensus, which resembles the segregated start scenario. Therefore, it takes longer for graphs with smaller sizes to break the local consensus and escape the so called "stuck" situation.

However, for the segregated start, it immediately starts with the "worst" case setting where the two communities have diverging opinions, so overall it takes longer to leave "stuck" situation compared with graphs of the same size in the "empty" start scenario. Additionally, graphs with larger sizes in the segregated setting more easily encounter a timeout. This may be because larger graphs occasionally time out even if they are not really "stuck" because they take longer to reach consensus in any event.

Theoretical Analysis. Next we will analyze the consensus time for the naming game on $Hom(G,p)$ where G has $2n$ agents divided into two equally sized disconnected communities with segregated start.

Theorem 4.1. *Let G be the disjoint union of two n cliques, each of size n. Then for the segregated start naming game on $Hom(G,p)$, there exists a constant $p_0 \approx 0.110$ such that if $0 \le p < p_0$ the expected consensus time is $\exp(\Omega(n))$.*

Here we sketch a proof of theorem. A full proof appears in the appendix.



(a)

(b)

Figure 4: Evaluations of reaching global consensus for different initialization scenarios and sizes of graphs. Row 1: Consensus time for (a) disjoint cliques (b) tree structure; row 2: fraction of nodes failing to reach consensus, based on different probability of random edges p for (a) disjoint cliques (b) tree structure; (c) normalized number of random edges used for communication as a function of the probability of random edges p.

To prove this theorem, we formulate the naming game as a nonhomogenous random walk on $\mathbb{Z}^4$ and relate this nonhomogenous random walk to a corresponding autonomous system in $\mathbb{R}^4$.

In the segregated start scenario, every node has an initial opinion, therefore no new name will be generated, and nodes inventory will be either $\{0\}, \{1\}$, or $\{0, 1\}$. Due to the symmetry among nodes, at each step t we only need to keep track of the number of nodes in the two groups whose inventory corresponds to the three categories above. Moreover, because the total number of two communities are n, we can use four variables to discribes this random process: fraction of $\{0\}, \{1\}$ nodes in two communities.

As the size of community increase, the above process is closely related to its mean field which can be seen as a autonomous system in $\mathbb{R}^4$. We show that this system has a stable fixed point as long as $0 \leq p < p_0$. To proof Theorem 4.1 we show two things with the help in autonomous system:

1. Global behaviour: the random process $X(t)$ will initially "converge" to a point corresponding to the stabile fixed point of the autonomous system.
2. Local behaviour: the random process $X(t)$ takes exponential time to leave the regions corresponding to the regions around stable fixed point of the autonomous system.

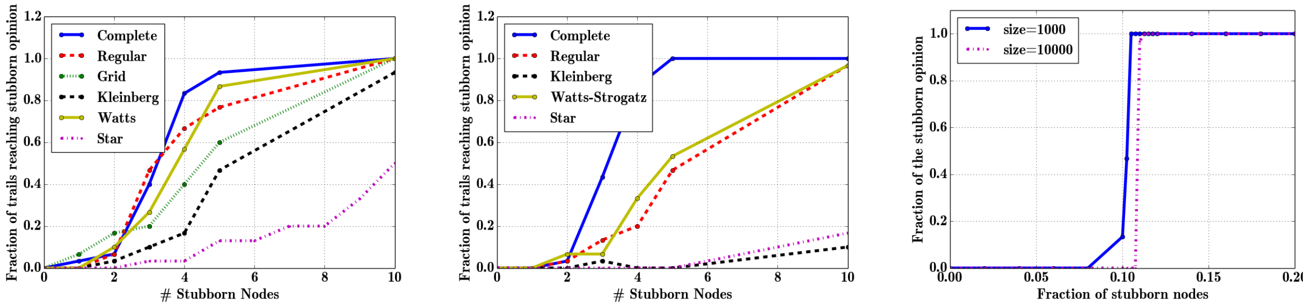
Tree Structure

In this section, we systematically study the Naming Game on trees and examine how the naming game converges when applying $\mathrm{Hom}(\cdot, p)$ to the tree structure. We show that convergence is substantially sped up for random edges added with small probabilities.

In Figure 4 (row 1 (b)) we evaluate the time to reach consensus based on the probability p of choosing random edges. It is clear that for trees with smaller depth (d) and more branches, the time to reach consensus is larger. Compared with Figure 4 (row 1 (a)) we see that by adding random edges, the tree graph is much less likely to encounter a time-out than the densely connected community graph. Figure 4 (row 2 (b)) show the fraction of agents failing to reach consensus as a function of p. Additionally, though the additional pairs can break up the sparse community structure and help to accelerate the converging process, redundant communications may be introduced at the same time. Therefore in Figure 4 (c) for various p, we present the total number of time the dynamics choose a homogenous edge before consensus is reached, normalized by time it takes homogeneous graph (clique) to reach consensus. We can see that there is actually an tipping point where the homogenous edges are used the least, which implies the edges of the original tree actually help towards consensus. Above this point, the homogeneous edges provide unnecessary communication redundancy.

5 Stubborn Nodes

In this section we introduce another aspect of asymmetry and heterogeneity. We introduce special agents called "stubborn" nodes, which never change their own opinions and aim to influence the whole network. The topic is also related to the robustness of the naming game in the real world setting, in which a small number of nodes can be malicious and



(a) (b) (c)

Figure 5: Evaluation for early stage coordinated stubborn nodes within different graphs. (a) fraction of trials converging to the stubborn nodes' opinion, as a function of the number of stubborn nodes with size 1000; (b) fraction of trials converging to the stubborn nodes' opinion, as a function of the number of stubborn nodes with size 10000; (c) fraction of nodes converging to the stubborn nodes' opinion in the late stage situation for complete graph of size 1000 and 10000.

not follow the protocol. The primary question we want to ask is: how and when can such nodes affect the opinion/name to which the dynamics converge? There are two important factors to consider here – the network topology and the time when the stubborn nodes are activated. Here we consider two situations: 1) the stubborn nodes join at the beginning of the game; 2) the nodes become "stubborn" after the graph has converged to one global opinion. Figure 5 (a) (b) show the fraction of trials converging to the stubborn nodes' preference based on the number of stubborn nodes in situation 1) for graphs of size 1000 and 10000, respectively. From Figure 5 (a), it is clear that in some graphs (e.g. the clique) even a small constant (e.g. 5) number of stubborn nodes can guarantee convergence to a particular name. Note that as the number of nodes increase, the curve barely changes, and if anything, becomes a sharper threshold. However, in others networks (e.g. star, grid, Kleinberg model), the number of required nodes seems to grow with the size of the graph.

This shows that in certain networks these protocols are not robust to rogue/stubborn agents. By comparing Figure 5 (a) and (b)), we see that the complete graph is not affected much by its size in terms of the influence efficiency of the "stubborn" nodes. However, in the Kleinberg and star graphs the number of stubborn nodes needed greatly depends on the size of the network. Note that here we choose the same number of stubborn nodes because complete, regular and Watts-Strogatz graphs actually perform similarly with size 1000 on these number of stubborn nodes.

Additionally, we show that in the complete graph, manipulating the name after convergence is much harder than before: there exists a value $p_0 \in [0, 1]$, such that if an adversary controls more than p_0 fraction of the node, consensus can be easily manipulated and otherwise it will take exponential time to manipulate the consensus. In Figure 5 (c) we verify this empirically by showing the fraction of trials converging to the stubborn nodes' preference (before timing out) based on the fraction of "stubborn" nodes within the network. It shows that at least 10% such stubborn nodes are needed to manipulate the opinion of the original graph empirically.

We provide theoretical analysis on the lower bound for the number of "stubborn" nodes required to manipulate the global consensus to align with the "stubborn" nodes' in a complete graph. In completed graph if the naming game converges to opinion 1, we want to answer the following question: what fraction of stubborn nodes with opinion 0 are required in order to convert the graph's consensus to 0 in polynomial time?

Theorem 5.1. *Given the naming game with p fraction of stubborn nodes defined above, there exists a constant $p_0 \approx 0.108$ such that for all $0 \leq p < p_0$ the expected consensus time is $\exp(\Omega(n))$. Additionally, if $p_0 < p \leq 1$ for all $\epsilon > 0$ the fraction of original opinion is smaller than ε after $O(n)$ steps.*

The proof appears in the appendix.

6 Conclusions

Our work sheds light on how asymmetric and heterogeneous agents, in both network topology and node types, affect the Naming Game with respect to convergence rate and the converged state. We show that in network structure complete asymmetry as in the star network is beneficial while asymmetry represented in community structures slow down global convergence. We rigorously studied how community structure can prohibit global convergence, and we also theoretically prove how much additional communication is required to facilitate convergence for graphs with community structures. Besides, we also analyze the robustness of different topologies (protocols) and discovered that only about 1% stubborn nodes can make a big difference in the behavior if they are introduced early in the game, while the introduction of stubborn nodes after convergence is achieved does not change much. Additionally, we theoretically prove how many stubborn nodes are needed to convert the global consensus for a complete graph, when they emerge after the convergence is reached. We believe these insights improve our foundational understanding of social dynamics in multi-agent systems and will spur on further insightful studies on this topic.

References

Backstrom, L.; Huttenlocher, D.; Kleinberg, J.; and Lan, X. 2006. Group formation in large social networks: membership, growth, and evolution. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 44–54. ACM.

Baronchelli, A.; Felici, M.; Loreto, V.; Caglioti, E.; and Steels, L. 2006a. Sharp transition towards shared vocabularies in multi-agent systems. *Journal of Statistical Mechanics: Theory and Experiment* 2006(06):P06014.

Baronchelli, A.; Loreto, V.; DallAsta, L.; and Barrat, A. 2006b. Bootstrapping communication in language games: Strategy, topology and all that. In *Proceedings of the 6th International Conference on the Evolution of Language*, volume 2006, 11–18. World Scientific Press.

Baronchelli, A.; Dall'Asta, L.; Barrat, A.; and Loreto, V. 2007. The role of topology on the dynamics of the naming game. *The European Physical Journal Special Topics* 143(1):233–235.

Baronchelli, A.; Loreto, V.; and Steels, L. 2008. In-depth analysis of the naming game dynamics: the homogeneous mixing case. *International Journal of Modern Physics C* 19(05):785–812.

Bollobás, B. 1998. Random graphs. In *Modern Graph Theory*. Springer. 215–252.

Brighton, H., and Kirby, S. 2001. The survival of the smallest: Stability conditions for the cultural evolution of compositional language. In *European Conference on Artificial Life*, 592–601. Springer.

Centola, D., and Baronchelli, A. 2015. The spontaneous emergence of conventions: An experimental study of cultural evolution. *Proceedings of the National Academy of Sciences* 112(7):1989–1994.

Dall'Asta, L.; Baronchelli, A.; Barrat, A.; and Loreto, V. 2006. Agreement dynamics on small-world networks. *EPL (Europhysics Letters)* 73(6):969.

Ellison, G. 2000. Basins of attraction, long-run stochastic stability, and the speed of step-by-step evolution. *The Review of Economic Studies* 67(1):17–45.

Franks, H.; Griffiths, N.; and Jhumka, A. 2013. Manipulating convention emergence using influencer agents. *Autonomous Agents and Multi-Agent Systems* 26(3):315–353.

Gong, T.; Ke, J.; Minett, J. W.; and Wang, W. S. 2004. A computational framework to simulate the coevolution of language and social structure. In *Artificial Life IX: Proceedings of the 9th International Conference on the Simulation and Synthesis of Living Systems*, 158–64.

Hurford, J. R. 1989. Biological evolution of the saussurean sign as a component of the language acquisition device. *Lingua* 77(2):187–222.

Kleinberg, J. 2000. The small-world phenomenon: An algorithmic perspective. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, 163–170. ACM.

Lu, Q.; Korniss, G.; and Szymanski, B. K. 2009. The naming game in social networks: community formation and consensus engineering. *Journal of Economic Interaction and Coordination* 4(2):221–235.

Nowak, M. A., and Krakauer, D. C. 1999. The evolution of language. *Proceedings of the National Academy of Sciences* 96(14):8028–8033.

Nowak, M. A.; Plotkin, J. B.; and Jansen, V. A. 2000. The evolution of syntactic communication. *Nature* 404(6777):495–498.

Nowak, M. A.; Plotkin, J. B.; and Krakauer, D. C. 1999. The evolutionary language game. *Journal of Theoretical Biology* 200(2):147–162.

Pickering, W., and Lim, C. 2016. Solution of the multistate voter model and application to strong neutrals in the naming game. *Physical Review E* 93(3):032318.

Steels, L., and McIntyre, A. 1998. Spatially distributed naming games. *Advances in complex systems* 1(04):301–323.

Steels, L. 1995. A self-organizing spatial vocabulary. *Artificial life* 2(3):319–332.

Steels, L. 2005. The emergence and evolution of linguistic structure: from lexical to grammatical communication systems. *Connection science* 17(3-4):213–230.

Strogatz, S. H. 2014. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Westview press.

Watts, D. J., and Strogatz, S. H. 1998. Collective dynamics of small-worldnetworks. *Nature* 393(6684):440–442.

Wormald, N. C. 1995. Differential equations for random processes and random graphs. *The annals of applied probability* 1217–1235.

A Preliminary

Our technique to prove Theorem 4.1 and Theorem 5.1 which combines mean field approximation and stability of differential systems. We think our technique will apply to other settings as well. At a very high level: first we relate the random process to a differential equation, and next we characterize the long term behaviour of differential equation.

Mean field approximation

There is extensive literature about stochastic processes and its mean field approximation e.g. (Ellison 2000). Given a nonhomogeneous random walk $X(t)$ in $\mathbb{Z}^\ell$ we can associate the behavior of it with the corresponding differential equation in $\mathbb{R}^\ell$. Formally, let $X_n(t)$ be a discrete time Markov chain on $\mathbb{Z}^\ell$ with parameter n which is time-homogeneous and the increments of the walk are uniformly bounded by β. As a result, random vectors $X_n(t+1) - X_n(t)$ have well defined moments, which depend on $X(t)$ and n. In particular, an important quantity is the one-step *mean drift vector* $F_n : \mathbb{R}^\ell \to \mathbb{R}^\ell$ defined to be

$$F_n(X) = E[X_n(t+1)X_n(t)|X_n(t) = X]. \quad (1)$$

In particular if there exists a function f independent of n such that $F_n(X) = f(\frac{X}{n})$, then there is a close relationship between X and the x which we define as a solution of the following autonomous differential system

$$x' = f(x) \quad (2)$$

with initial condition $x(0) = X(0)/n$.

The following theorem shows that the differential equation approximates the original random walk $X(t)$ such that $X(t) \approx n\hat{x}(\frac{t}{n})$ under proper conditions.

Theorem A.1 (Wormald's method (Wormald 1995)). *For $1 \leq \ell \leq a$ where a is fixed, let $y_\ell : S^{(n)+} \to \mathbb{R}$ and $f_\ell : \mathbb{R}^{a+1} \to \mathbb{R}$ such that for some constant C_0 and all $\ell, |y_\ell(h_t)| < C_0 n$ for all $h_t \in S^{(n)+}$ and n. Let $Y_\ell(t)$ denote the random counterpart of $y_\ell(h_t)$. Assume the following three conditions hold:*

1. *(Boundedness) For some functions $\beta = \beta(n) \geq 1$ and $\gamma = \gamma(n)$, the probability that*
$$\max_\ell |Y_\ell(t+1) - Y_\ell(t)| \leq \beta$$
conditional upon H_t, is at least $1 - \gamma$ for $t < T_D$.
2. *(Trend) For some function $\lambda_1 = \lambda_1(n) = o(1)$, for all $\ell \leq a$*
$$|\mathbb{E}[Y_\ell(t+1) - Y_\ell(t)|H_t] - f_\ell(\frac{t}{n}, \frac{Y_1(t)}{n}, ..., \frac{Y_a(t)}{n}) \leq \lambda_1$$
for $t \leq T_D$.
3. *(Lipschitz) Each function f_ℓ is continuous, and satisfies a Lipschitz condition, on*
$$D \cap \{(t, z_1, ..., z_a) : t \geq 0\},$$
with the same Lipschitz constant for each ℓ.

Then the following are true.

1. *For $(0, \hat{z}_1, ..., \hat{z}_a) \in D$ the system of differential equations*
$$\frac{dz_\ell}{dx} - f_\ell(x, z_1, .., z_a), \ell = 1, ..., z \quad (3)$$
have a unique solution in D for $z_\ell : \mathbb{R} \to \mathbb{R}$ passing through $z_\ell(0) = \hat{z}_\ell$ for $1 \leq \ell \leq a$, which extends to points arbitrarily close to the boundary of D;
2. *Let $\lambda > \lambda_1 + C_0 n\gamma$ with $\lambda = o(1)$. For a sufficiently large constant C with probability $1 - O(n\gamma + \frac{\beta}{\lambda}\exp(\frac{-n\lambda^3}{\beta^3}))$,*
$$Y_\ell(t) = nz_\ell(\frac{t}{n}) + O(\lambda n) \quad (4)$$
uniformly for $0 \leq t \leq \sigma n$ and for each ℓ where $z_\ell(x)$ is the solution in Equation (3) with $\hat{z}_\ell = \frac{Y_\ell(t)}{n}$, and $\sigma = \sigma(n)$ is the supremum of those x to which the solution can be extended before reaching within ℓ^∞-distance $C\lambda$ of the boundary of D.

Stability of autonomous system

Stability capture the long term behaviour of (2). Here is some notation: a point $\bar{x} \in \mathbb{R}^\ell$ is called an *equilibrium point* of system (2) if $f(\bar{x}) = 0$. Moreover the equilibrium $\bar{x}$ is *asymptotically stable* if $\forall\epsilon > 0, \exists\delta > 0$ such that $||x(0) - \bar{x}|| \leq \delta \Rightarrow ||x(t) - \bar{x}|| \leq \epsilon, \forall t$ and $\exists\delta > 0$ such that $\lim_{t\to\infty} ||x(t) - \bar{x}|| = 0$. The stability of the system can be determined by the linearization of the system which is stated below.

Theorem A.2 (Lyapunov's indirect method (Strogatz 2014)). *Let x^* be an equilibrium point for $x' = f(x)$ where $f : \mathcal{D} \to \mathbb{R}^d$ is continuously differentiable and $\mathcal{D}$ is a neighborhood of x^*. Let $A = \frac{\partial f}{\partial x}|_{x=x^*}$ then x^* is asymptotically stable if A is* Hurwitz*, that is $Re(\lambda_i) < 0$ for all eigenvalues of A.*

Moreover, there exists an close set $U \subseteq \mathcal{D}$ and $x^* \in U$ and a potential function $V : U \to \mathbb{R}$ such that $V(x^*) = 0$, and $V(x) > 0$, $\frac{d}{dt}(V(x)) < 0$ for $x \in U \setminus x^*$.

However the above theorem only captures the behaviour of x when it is close enough to the stabile fixed point x^*. On the other hand for global stability, the following theorems is quite useful when system (2) is in the plane. To state the theorem we need to introduce more terminology. A set is bounded if it is contained in some cycle $\{x \in \mathbb{R}^2 | ||x - \alpha < C\}$ for some $\alpha \in \mathbb{R}^2$ and $C > 0$. A point $p \in \mathbb{R}^2$ is called an ω-limit point of the trajectory $\Gamma_{z_0} = \{z(t)|t \geq 0, z(0) = z_0\}$ of the system (2) if there is a sequence $t_n \to \infty$ such that $\lim_{n\to\infty} x(t_n) = p$.

Theorem A.3 (Poincare-Bendixon Theorem (Strogatz 2014)). *Let $z' = H(z)$ be a system of differential equations defined on E an open subset in $\mathbb{R}^2$ where H is differentiable. Suppose a forward orbit with initial condition z_0 $\Gamma_{z_0} = \{z(t)|t \geq 0, z(0) = z_0\}$ is bounded. Then either*

- *$\omega(z_0)$ contains a fixed point*
- *$\omega(z_0)$ is a periodic orbit*

The following theorem gives us a sufficient condition for nonexistence of periodic orbit

Theorem A.4 (Bendixson's Criteria (Strogatz 2014))**.** *Let H be differentiable in E where E is a simply connected region in $\mathbb{R}^2$. If the divergence of the vector field H is not identically zero and does not change sign in E then $z'H(x)$ has no closed periodic orbit lying entirely in E.*

Note that the theorem only holds for two dimensions system and fails in general.

B Main Results

The main idea used to prove both Theorem 4.1 and Theorem 5.1 is to show the existence of a stable fixed point x^* of the solution to differential system (2) and then to relate this stable fixed point to the nonhomogeneous random walk (1) by showing:

1. Global behaviour: the random process $X(t)$ will initially "converge" to a point corresponding to the stabile fixed point of the autonomous system.
2. Local behaviour: random process $X(t)$ takes exponential time to leave the region corresponding to a regions around stabile fixed point of the autonomous system.

Here we prove a auxiliary theorem for the second part.

Theorem B.1. *If x^* is an asymptotically stable equilibrium of* (2)*, given a closed set U containing x^* there exists $r_a > 0$ such that in system* (1) *if $||X(t_0)/n - x^*|| \leq r_a$ then*

$$\mathbb{E}[\arg\min_{\tau > t_0}\{X(\tau) \notin U\}\big| ||X(t_0)/n - x^*|| \leq r_a] = \exp(\Omega(n)).$$

To prove Lemma B.1, we use the second part of Lyapunov's indirect method Theorem A.2, which shows the existence of a potential function $\hat{V}(x)$ at some region around the asymptotically stable fixed point in system (2) such that the value of potential function is strictly decrease along the trajectory. On the other hand, the counterpart of that potential function in (1) will be a supermartingale $V(X(t))$ and we use the optional stopping time to show that it takes an exponential time for the supermartingale to increase by constant.

Proof of Lemma B.1. By Theorem A.2, we know that there exists a potential function V and an open region $U \subseteq \mathcal{D}$ such that $V(x^*) = 0$, and $V(x) > 0$, $\frac{d}{dt}(V(x)) < 0$ for $x \in U \setminus x^*$. Now we consider a random process

$$W(i) = V\left(\frac{X(i)}{n}\right)$$

and the conditional expectation is

$$\begin{aligned}
&\mathbb{E}[W(i+1) - W(i)|X(i)] \\
=&\mathbb{E}[V(\frac{X(i+1)}{n}) - V(\frac{X(i)}{n})|X(t)] \\
=&\nabla V(\frac{X(i)}{n}) \cdot (\frac{\mathbb{E}[X(i+1) - X(i)|X(i)]}{n}) + O(\frac{1}{n^2}) \\
=&\nabla V(\frac{X(i)}{n}) \cdot \frac{f(\frac{X(i)}{n})}{n} + O(\frac{1}{n^2}) \\
=&\frac{1}{n}\frac{d}{dt}V(x)\Big|_{x=\frac{X(i)}{n}} + O(\frac{1}{n^2})
\end{aligned} \tag{5}$$

Therefore $W(i)$ is a supermartingale such that $\mathbb{E}[W(i+1) - W(i)|X(i)] < 0$ when $\frac{X(t)}{n} \in U \setminus \{x^*\}$ and n is large enough.

The idea is to use the optional stopping theorem by proving the process $X(t)$ is not likely to pass through the annulus $B_{r_b} \setminus B_{r_a}$ for some properly choosen r_a, r_b. Here we need to use the properties of the potential function V from Theorem A.2. Note that U is open, there exists $r_b > 0$ such that a open set $B_{r_b} = \{||x - x^*|| < r_b\} \subseteq U$. Because the boundary $U \setminus B_{r_b}$ is compact and V is continuous, there exists $\min_{x \in B_{r_b}} V(x)$ which is denoted as l_b. On the other hand, because $V(x^*) = 0$ and V is continuous, there exists a close set $\bar{B_{r_a}}$ where $0 < r_a < r_b$ such that $l_a = \max_{x \in \bar{B_{r_a}}} V(x) \leq 0.3 l_b$.

Given such r_a, r_b if $X(t_0)/n \in B_{r_a}$ at some time t_0 and the system leaves the stable region U at time $t_1 > t_0$ there exists σ,τ when n is large enough such that

$$\tau = \arg\min_{t_0 < t < t_1}\{X(t)/n \in U \setminus B_{r_b}\}$$

$$\sigma = \arg\max_{t_0 < s < \tau}\{X(s)/n \in B_{r_a}\}$$

which gives us

$$W(\sigma) < 0.5 l_a, \text{and } W(\tau) \geq l_b$$

Moreover by the definition of σ, τ, for all $\sigma \leq t < \tau$ the random process $X(t)$ would stay in the annulus $B_{r_b} \setminus B_{r_a}$. Therefore for all t such that $\sigma \leq t < \tau$, we have $W(t)$ is a strict supermartingale

$$W(t) = \frac{1}{n}\frac{d}{dt}V(x)\Big|_{x=\frac{X(i)}{n}} + O(\frac{1}{n^2}) = \frac{-h}{2n} < 0$$

where constant $-h = \max_{x \in B_{r_b} \setminus B_{r_a}} \frac{d}{dt}V(x)\Big|_x < 0$ since the annulus is compact.

Therefore by standard optional stopping time theorem with initial state $W(\sigma + 1)$ where $l_a < W(\sigma + 1) < l_b$ the average time for $W(t)$ to hit $W(t) \geq l_b$ is $\exp(\Omega(hn)) = \exp(\Omega(n))$. ☐

C Proof of Theorem 4.1

Recall that we want to formulate the naming game as nonhomogenous random walk on $\mathbb{Z}^4$ and relate this nonhomogenous random walk to a correpsonding autonomous system in $\mathbb{R}^4$ to study consensus time. Note that we can use four variables to describe this random process: fraction of $\{0\}, \{1\}$ nodes in two communities by following notations.

A_t	community1	community2
$\{0\}$	$R_1(t)$	$R_2(t)$
$\{1\}$	$B_1(t)$	$B_2(t)$
$\{0,1\}$	$M_1(t)$	$M_2(t)$

Since $n = R_1(t) + B_1(t) + M_1(t) = R_2(t) + B_2(t) + M_2(t)$ for all t, it's sufficient to consider $X(t) = (R_1(t), B_1(t), R_2(t), B_2(t))$ in $\mathbb{Z}^4$ with initial state $X(0) = (n, 0, 0, n)$ and the naming game reaches consensus at T when $X(T) = (n, 0, n, 0)$ or $(0, n, 0, n)$.

We can now define $F(\cdot)$ as the mean field of this system (as in Equation (1)):

$$F(X(t)) = \mathbb{E}[X(t+1) - X(t)|X(t)]. \quad (6)$$

Our approach to understand the behavior of X is mainly inspired by the stability property of nonlinear autonomous systems. We define $f(\cdot)$ such that $F_n(X) = f(\frac{X}{n})$ and then we can relate the nonhomogeneous random walk X to the solution of $x' = f(x)$ as in (2).

Intuitively we will prove that there exists p such that the system has an "undesirable" asymptotically stable points x^* (which will be defined mathematically in appendix)

$$x^* = (r^*, b^*, b^*, r^*)$$

where $r^* = \frac{e^2+\sqrt{-4e+6e^2-e^4}}{2e}$, $b^* = \frac{e^2-\sqrt{-4e+6e^2-e^4}}{2e}$ and $p = \frac{2}{3}(1-e)$ such that the random process $X(t)$ in Equation (6) will

1. Reach some region of nx^*.
2. Given $X(T_0)$ is in some region of nx^* the expected consensus time of the corresponding naming game is exponential in the size of each group $\exp(\Omega(n))$.

These two conclusions can be proved by the following two lemmas, respectively and the proof of Theorem 4.1 follows directly from the above two Lemmas.

Lemma C.1. *Given the naming game defined above, if* $0 \le p < \frac{4-2\sqrt{3}}{3} \approx 0.178$ *given arbitrary constant* $r_a > 0$ *the random walk* $X(t)$ *will converge to* x^**. That is there exist* $T_0 = O(n)$ *such that* $||X(T_0)/n - x^*|| \le r_a$ *with probability* $1 - O(\frac{\log n}{\exp(\frac{-n}{\log^3 n})})$

Lemma C.2. *Given the naming game defined above, there exists a constant* $p_0 \approx 0.110$ *such that for all* $0 \le p < p_0$ *there exists some constant* $r_a > 0$ *such that if* $||X(T_0)/n - x^*|| \le r_a$ *then the consensus time is* $\exp(\Omega(n))$

Now we need to quantify the evolution of this process. Recalled that our naming game defined in (6)

$$\begin{aligned}
&\mathbb{E}[R_1(t+1) - R_1(t)|X(t)] = \frac{1}{2}\{(1 - \frac{R_1}{n} - 2\frac{B_1}{n} + (\frac{B_1}{n})^2)\\
&+ \frac{p}{2}(\frac{-R_1}{2n} + \frac{B_1}{n} + \frac{R_2}{2n} - \frac{B_2}{n} - (\frac{B_1}{n})^2 - \frac{3R_1B_2}{2n^2} - \frac{B_1R_2}{2n^2})\}\\
&\mathbb{E}[B_1(t+1) - B_1(t)|X(t)] = \frac{1}{2}\{(1 - \frac{B_1}{n} - 2\frac{R_1}{n} + (\frac{R_1}{n})^2)\\
&+ \frac{p}{2}(\frac{-B_1}{2n} + \frac{R_1}{n} + \frac{B_2}{2n} - \frac{R_2}{n} - (\frac{R_1}{n})^2 - \frac{3B_1R_2}{2n^2} - \frac{R_1B_2}{2n^2})\}\\
&\mathbb{E}[R_2(t+1) - R_2(t)|X(t)] = \frac{1}{2}\{(1 - \frac{R_2}{n} - 2\frac{B_2}{n} + (\frac{B_2}{n})^2)\\
&+ \frac{p}{2}(\frac{-R_2}{2n} + \frac{B_2}{n} + \frac{R_1}{2n} - \frac{B_1}{n} - (\frac{B_2}{n})^2 - \frac{3R_2B_1}{2n^2} - \frac{B_2R_1}{2n^2})\}\\
&\mathbb{E}[B_2(t+1) - B_2(t)|X(t)] = \frac{1}{2}\{(1 - \frac{B_2}{n} - 2\frac{R_2}{n} + (\frac{R_2}{n})^2)\\
&+ \frac{p}{2}(\frac{-B_2}{2n} + \frac{R_2}{n} + \frac{B_1}{2n} - \frac{R_1}{n} - (\frac{R_2}{n})^2 - \frac{3B_2R_1}{2n^2} - \frac{R_2B_1}{2n^2})\}\\
&R_1(0) = n, B_1(0) = 0, R_2(0) = 0, B_2(0) = n
\end{aligned}$$

has corresponding autonomous differential system as follow.

$$\begin{aligned}
r_1' &= \frac{1}{2}\{(1 - r_1 - 2b_1 + b_1^2\\
&+ \frac{p}{2}(\frac{-1}{2}r_1 + b_1 + \frac{1}{2}r_2 - b_2 - b_1^2 - \frac{3}{2}r_1b_2 - \frac{1}{2}b_1r_2)\}\\
b_1' &= \frac{1}{2}\{(1 - b_1 - 2r_1 + r_1^2\\
&+ \frac{p}{2}(\frac{-1}{2}b_1 + r_1 + \frac{1}{2}b_2 - r_2 - r_1^2 - \frac{3}{2}b_1r_2 - \frac{1}{2}r_1b_2)\}\\
r_2' &= \frac{1}{2}\{(1 - r_2 - 2b_2 + b_2^2\\
&+ \frac{p}{2}(\frac{-1}{2}r_2 + b_2 + \frac{1}{2}r_1 - b_1 - b_2^2 - \frac{3}{2}r_2b_1 - \frac{1}{2}b_2r_1)\}\\
b_2' &= \frac{1}{2}\{(1 - b_2 - 2r_2 + r_2^2\\
&+ \frac{p}{2}(\frac{-1}{2}b_2 + r_2 + \frac{1}{2}b_1 - r_1 - r_2^2 - \frac{3}{2}b_2r_1 - \frac{1}{2}r_2b_1)\}\\
&r_1(0) = 1, b_1(0) = 0, r_2(0) = 0, b_2(0) = 1 \qquad (7)
\end{aligned}$$

Proof of Lemma C.2

With Theorem B.1, to prove Lemma C.2, it is sufficient to prove x^* is a stable fixed point.

Proof of Lemma C.2. With Theorem A.2, it is sufficient to show all the eigenvalues of $A = \frac{\partial f}{\partial x}|_{x=x^*}$ are negative. By elementary computation, the eigenvalues of A are

$$\frac{-e-5}{6} - D_1, \frac{-e-5}{6} + D_1$$
$$\frac{e^2-3}{2} - D_2, \frac{e^2-3}{2} + D_2$$

where $p = \frac{2}{3}(1-e)$ and

$$D_1 = \frac{1}{6}\sqrt{\frac{(1-e)(-8e^4 - 36e^3 + 7e^2 + 153e + 64)}{e}}$$
$$D_2 = \frac{1}{2}\sqrt{(1-e)(-e^3 - 5e^2 + e + 25)}$$

Therefore A is Hurwitz and x^* is asymptotically stable if $e > 0.835$ and $0 \le p < 0.110$ ☐

Proof of Lemma C.1

To proof Lemma C.1 we prove two claims:

1. The solution x to the differential equation in (7) converges to x^*;
2. the limit behavior of random process in (6) can be approximated by x in (7), that is $\lim_{n\to\infty} X(nt)/n \approx x(t)$.

With these two claims we can conclude given any $r_a > 0$ there exists t_0 such that $||X(t)/n - x^*|| < r_a$ for all $t > t_0$ with high probability. For the first claim we use Poincare-Bendixon Theorem A.3 and use Wormald's differential equation method A.1 to prove the second.

Proof of Lemma C.1. First, by the symmetry of the system and initial conditions $\hat{r}_1 = \hat{b}_2 = 1$ and $\hat{b}_1 = \hat{r}_2 = 0$. we can assume that $\hat{r}_1(t) = \hat{b}_2(t)$ and $\hat{b}_1(t) = \hat{r}_2$ for all $t \geq 0$, and the system of differential equations is equivalent to the following

$$r' = (1 - r - 2b + b^2) + \frac{1-e}{2}(b - r - b^2 - r^2)$$
$$b' = (1 - b - 2r + r^2) + \frac{1-e}{2}(r - b - r^2 - b^2)$$
$$\text{where } r(0) = 1, \text{ and } b(0) = 0$$

where $r(t) = \hat{r}_1(t) = \hat{b}_2(t)$ $b(t) = \hat{r}_1(t) = \hat{b}_2(t)$ and $p = \frac{1-e}{3}$, and the system will have stable fixed point $r^* = \frac{e^2+\sqrt{-4e+6e^2-e^4}}{2e}$ and $b^* = \frac{e^2-\sqrt{-4e+6e^2-e^4}}{2e}$, and we take

$$x^* = (r^*, b^*, b^*, r^*)$$

Note that such x^* exists if $-4 + 6e - e^3 \geq 0$, i.e. $0 \leq p \leq \frac{4-2\sqrt{3}}{3} \approx 0.178$.

To apply Theorem A.3 in (8), we need to show the orbit of (r, b) is bounded and there is no periodic cycle. It is easy to see that $r(t), b(t)$ are bounded in interval $[0, 1]$. Moreover because if $r(t) = b(t)$ for some t then $r(t') = b(t')$ for all $t' \geq t$, we have $r(t) \geq b(t)$. Combining these two observations we have (r, b) is bounded in $\Omega = \{(r, b)|r \geq b, 0 \leq r, b \leq 1\}$. On the other hand, because $\nabla \cdot H = -2 + \frac{1-e}{2}(-2 - 2r - 2b) < 0 \forall (r, b) \in \Omega$ which, by Theorem A.4, proves there is no closed orbit. Therefore we have proven the first claim: $\lim_{t\to\infty}(r(t), b(r)) = (r^*, b^*)$ by Theorem A.3. Furthermore in (7) we have

$$||x(t) - x^*|| < 0.5 r_a \forall t > t_0. \tag{8}$$

For the second claim, we want to show the original process in (6) can be approximated by (7). It is not hard to show that the process is bounded by $\beta = 1$ and $\gamma = 0$, and by taking $\lambda = O(\frac{1}{\log(n)})$ we have with probability $1 - O(\log n \exp(-\frac{n}{\log^3 n})$

$$X(nt)/n = x(t) + O(\frac{1}{\log n}) \tag{9}$$

in terms of each component.

Combining (8) and (9) we have with probability $1 - O(\log n \exp(-\frac{n}{\log^3 n}))$

$$||X(nt)/n - x^*|| \leq r_a, \forall t > t_0$$

when n is large enough. □

D Proof of Theorem 5.1

We define stubborn node which has different behavior in naming game. A node s is stubborn if its inventory will not change the process $A_t(s) = A_0(s)$ even when it is the speaker or listener, and we call node s is *stubborn node* with $A_0(s)$, and we call other node as *ordinary nodes*. Here we consider that on completed graph if the naming game is already consensus on opinion 1. The Theorem 5.1 gives a way to understand the following question: how many nodes stubborn with opinion 0 do we make in order to change the graph consensus on opinion 0 in polynomial time?

Theorem D.1 (Restate theorem 5.1)**.** *Given the naming game with p fraction of stubborn nodes defined above, there exists a constant* $p_0 \approx 0.108$ *such that for all* $0 \leq p < p_0$ *the expected consensus time is* $\exp(\Omega(n))$. *Additionally, if* $p_0 < p \leq 1$ *for all* $\epsilon > 0$ *the fraction of original opinion is smaller than* ε *after* $O(n)$ *steps.*

Similar to the proof of theorem 4.1, we formulate this process as nonhomogenous random walk on $\mathbb{Z}^2$ and relate this nonhomogenous random walk to a correpsonding autonomous system in $\mathbb{R}^2$ to study consensus time.

Model Description

Given a completed graph G which has n nodes and the weight of every pair of node is uniform, if every nodes consensus on 1, we want to make p fraction of nodes stubborn on 0, and all the set of stubborn nodes S such that $|S| = pn$. That is $\forall s \in S, A_0(s) = \{0\}$ and for all ordinary node $v \in V(G) \setminus S, A_0(v) = \{0\}$.

Because the symmetry of the completed graph, only the number of stubborn nodes matters, and we apply the same method in theorem 4.1 to simplify the notations. At time t, we define $X(t) = (R(t), B(t))$ as our state of Markov chain where $R(t)$ the number of ordinary node with inventory $\{0\}$, $B(t)$ the number of ordinary node with inventory $\{1\}$ and $M(t)$ be the number of ordinary node with inventory $\{0, 1\}$. Moreover we use $\tilde{n}$ to denote the number of ordinary nodes, $\tilde{n} = |V(G) \setminus S| = (1 - p)n$. Here we have

$$\mathbb{E}[R(t + 1) - R(t)|X(t)]$$
$$=(1 - p)^2(\frac{R}{\tilde{n}}\frac{M}{\tilde{n}} + (\frac{M}{\tilde{n}})^2 - \frac{R}{\tilde{n}}\frac{B}{\tilde{n}}) + p(1 - p)\frac{3}{2}\frac{M}{\tilde{n}}$$
$$\mathbb{E}[B(t + 1) - B(t)|X(t)]$$
$$=(1 - p)^2(\frac{B}{\tilde{n}}\frac{M}{\tilde{n}} + (\frac{M}{\tilde{n}})^2 - \frac{R}{\tilde{n}}\frac{B}{\tilde{n}}) - p(1 - p)\frac{B}{\tilde{n}}$$

and the corresponding autonomous differential system is

$$r' =(1 - p)^2(rm + m^2 - rb) + p(1 - p)\frac{3}{2}m$$
$$b' =(1 - p)^2(bm + m^2 - rb) - p(1 - p)b$$

Proofs

Similar to theorem 4.1, when $p < 0.108$ it is striaghtforward to show there exists a stable fixed point $x^* \neq (1, 0)$ and derived the following two lemmas to prove the first part of Theorem 5.1.

Lemma D.1. *Given the naming game defined above, there exists* $p_0 \approx 0.108$ *such that for all constant* $0 \leq p < p_0$ *there exists some constant* $r_a > 0$ *such that if* $||X(T_0)/n - x^*|| \leq r_a$ *then the consensus time is* $\exp(\Omega(n))$

Lemma D.2. *Given the naming game defined above, if constant* $0 \leq p < 0.108$ *given arbitrary constant* $r_a > 0$ *the random walk* $X(t)$ *will converge to* x^*. *That is there exists* $T_0 = O(n)$ *such that* $||X(T_0)/n - x^*|| \leq r_a$ *with probability* $1 - O(\frac{\log n}{\exp(\frac{-n}{\log^3 n})})$

For the second part of Theorem 5.1, since if $p > p_0$ the consensus point,$c^* = (1, 0)$ is the only fixed point of the system, we can use similar technique in Lemma C.1 and Theorem B.1 to prove given arbitrary small constant $\epsilon > 0$, $b(t) \leq \epsilon$ for $t = O(n)$.