Individual heterogeneity generating explosive system network dynamics

Pedro D. Manrique and Neil F. Johnson
Physics Department, University of Miami, Coral Gables, Florida 33126, USA

(Received 16 December 2017; published 21 March 2018)

Individual heterogeneity is a key characteristic of many real-world systems, from organisms to humans. However, its role in determining the system's collective dynamics is not well understood. Here we study how individual heterogeneity impacts the system network dynamics by comparing linking mechanisms that favor similar or dissimilar individuals. We find that this heterogeneity-based evolution drives an unconventional form of explosive network behavior, and it dictates how a polarized population moves toward consensus. Our model shows good agreement with data from both biological and social science domains. We conclude that individual heterogeneity likely plays a key role in the collective development of real-world networks and communities, and it cannot be ignored.

DOI: 10.1103/PhysRevE.97.032311

I. INTRODUCTION

The mechanisms by which single entities (e.g., molecules, cells, people, etc.) come together as groups underlie myriad physical, biological, and social processes [1–9]. Though multiple models have been proposed in order to explain the different aspects of real-world grouping behavior, the feature of heterogeneity among individuals has hardly been considered. Indeed, the universality of some collective processes across contrasting systems seems to have overshadowed the reality of intrinsically heterogeneous populations, and more importantly system-level evolution driven by individual heterogeneity. Percolation refers to the dynamical transition to large-scale connectivity by the addition of individual links [10,11]. Before the transition, all clusters are of negligible size compared to that of the whole system. From the transition point onward, the largest cluster gathers a finite non-negligible fraction of the total number of individuals. The simplest way to form and grow clusters is with the classic Erdós-Rényi (ER) model [12,13], in which at each time step a new link is established between two randomly selected nodes. This model reaches the percolation transition when the number of added links reaches half of the total number of nodes. At this point, the second moment associated with the cluster-size distribution presents a singularity and the system passes to a gel phase [14–16].

Careful recent work has been dedicated to exploring the dynamical control of percolation, particularly the transition point [17–19]. Depending on the system, researchers look for ways to either delay or accelerate the appearance of the percolation transition. This is made possible by introducing the element of competition among the nodes to be associated together. Instead of randomly selecting one pair of nodes to be linked, three or more nodes are selected to compete for link addition [18,19]. The rules for addition vary from model to model, however the direct aspect that is being explored is how the different potential new links might affect the size of the resultant new cluster. Thus, rules have been proposed so that only small clusters are formed, which delays the transition, or large clusters join, resulting in the contrary effect. These new models of large-scale connectivity show abrupt variations in the size of the largest cluster and hence are known as *explosive* percolation. Examples from the biological and social domain have been shown to have features akin to explosive percolation [20,21,26]. However, it is not obvious why these models should be able to ignore individuals' heterogeneity when determining which clusters are formed and when. Even in our everyday lives, we do not arbitrarily form clusters without some underlying factor determined by the characteristics of the individuals involved. For example, in social gatherings we often join with family members who by definition have similar genes to ours, while in a sport we join those having complementary skills in order to form a strong team. Our findings suggest that heterogeneity-based cluster formation and hence node-to-node affinity interactions can play a crucial role in generating explosive phenomena in real-world systems, and influence the point of transition. As a result, our work provides fresh insights into how the diversity of individuals could affect the overall dynamics of the system. This, in turn, connects to recent findings in the stability of synchronous states in oscillator networks, which indicate that these particular states can only be attained by nonidentical oscillators [22–24]. Moreover, the output of our models closely captures the dynamics of real-world systems such as online extremism and protein homology networks.

II. MODEL

The heterogeneity is introduced as a hidden variable x_i that we call the "character" and is assigned randomly to each node i from a distribution $q(x)$ [27,28]. For simplicity, we consider all $\{x\}$ to be real numbers between 0 and 1 since larger ranges can be easily rescaled to be within this interval. Also, we consider that the character values are constant in time, though this can be modified to include variations over time to account for experience or external influence. The mechanisms of link addition follow directly from the relationship among the $\{x\}$ values associated with the nodes to be linked. This is quantified by the similarity S_{ij} between node i and node j, which is defined as $S_{ij} = 1 - |x_i - x_j|$. Thus highly similar

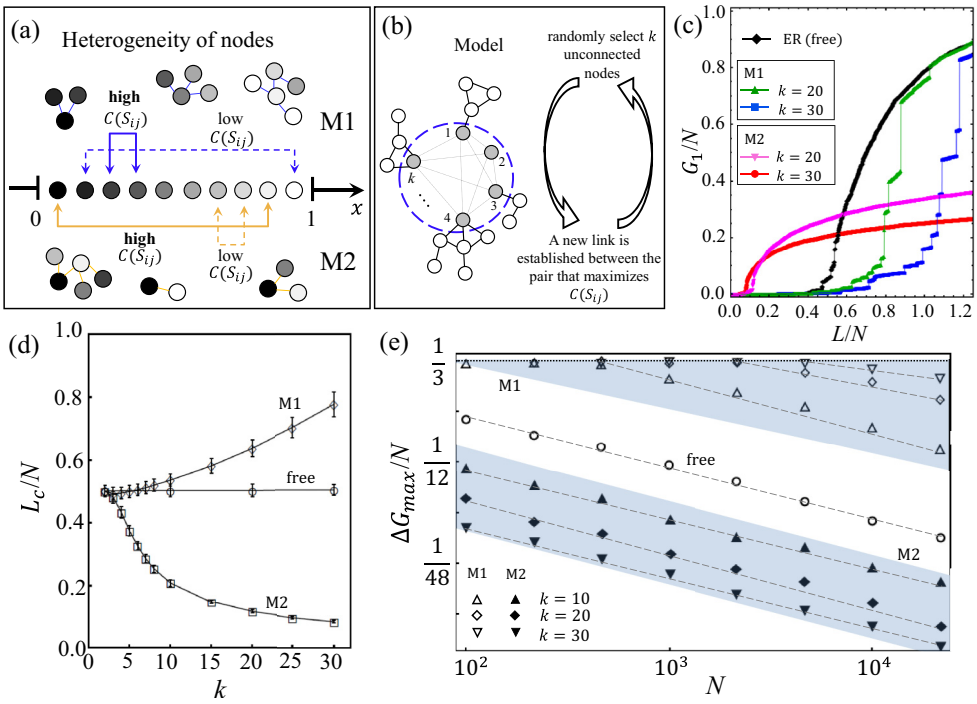

FIG. 1. Modeling heterogeneous grouping. (a) The heterogeneity of nodes is modeled by the character $\{x\}$, which is a real number defined in the [0,1] interval and randomly assigned to each node from a distribution $q(x)$. Our model of interacting nodes comprises two mechanisms: M1, which favors linking similar nodes (i.e., close in character), and M2, favoring dissimilar nodes (i.e., distant in character). (b) The model follows a two-step cycle where it first randomly selects a sample of k unconnected nodes and then selects the pair that maximizes the coalescence function $\mathcal{C}(S_{i,j})$ that links together. (c) Dynamics of the largest cluster G_1 as a function of the number of links for both heterogeneous formation mechanisms as well as the random case (i.e., free) and different values of sampling size k. The total population is $N = 10^4$ nodes. (d) Variation of the transition point for M1, M2, and the free model with the sampling size k computed at the instant when G_1/N and its growth rate exceed the value of the random process at $L/N = 0.5$. (e) Dependence of the largest gap in the growth of G_1 for both mechanisms (shadows) and the free model, and several values of sampling. Lines are guides to the eye. For (d) and (e) we show an average over 500 realizations.

nodes are close to each other in the x spectra and otherwise for highly dissimilar nodes. The mechanisms of link addition rest in the definition of the coalescence function $\mathcal{C}(S_{ij})$. We consider two complementary mechanisms for link addition: mechanism 1 (M1), favoring similar nodes, and mechanism 2 (M2), favoring dissimilar nodes [see Fig. 1(a)]. A system following M1 tends to form groups of alike individuals (e.g., kin) while M2 tends to form groups with unlike or complementary individuals (e.g., teams). Hence, for M1 a coalescence function is defined as $\mathcal{C}(S_{ij}) = S_{ij}$ while for M2, $\mathcal{C}(S_{ij}) = 1 - S_{ij}$.

Our model starts with all N nodes unconnected. At each time step a sample from the system is randomly selected and a new link is established between the pair of nodes that maximizes the coalescence function $\mathcal{C}(S_{ij})$ [see Fig. 1(b)]. Thus, the distribution $q(x)$, the sampling method and size, and the specific mechanism of link addition (i.e., M1 or M2) dictate the evolution of the system. Though the sampling can be either of nodes or links, the evolution of the network presents similar properties. Sampling of nodes can be related to individuals that ran into each other in the course of a time period (e.g., a single day), and among all the interactions only a few connections are established based on their mutual affinity. Here we will consider sampling of nodes for simplicity, with the link version in the supplemental material (SM) [25]. The smallest sampling size refers to two randomly selected nodes, which sets the limit of a random graph model (ER model). Note that this is independent of the link addition mechanism and the distribution $q(x)$. In a similar way, considering a distribution $q(x)$ to be a Dirac δ implies that all the nodes are identical (i.e., character-free) and hence the limit of a random graph model is also found independently of the remaining parameters. The former and latter observations refer to the competitive and diverse aspects of the network formation process, respectively. Unless both aspects are present, the system will follow an ER process.

The evolution of the system is typically described by means of the size of the largest cluster, which we call G_1. All present clusters sizes are denoted G_i, where i is the size rank starting from the largest (i.e., $i = 1$). Figure 1(c) shows the evolution of G_1 for several sampling sizes k, and mechanisms M1 and M2 when the character values are assigned randomly from a uniform distribution $q(x)$. The aggregation mechanisms and sampling sizes have contrasting macroscopic effects in the evolution of G_1. For example, for M1 the percolation transition shifts from the random case (character-free) point at $L/N = 0.5$ to greater values of L/N and exhibits macroscopic features akin to explosive percolation such as large jumps. On the other hand, for M2 the transition shifts to smaller values of L/N than the random case, and the growth tends to be smoother. In all cases, the shift tends to be more severe as the sampling size increases, as shown in Fig. 1(d).

These features can be explained by means of the formation mechanisms. M1 links pairs of nodes whose x values are the closest to each other. The probability density function (PDF) $f(y)$ of the similarity $y = S_{i,j}$ associated with the uniform distribution is $f(y) = 2y$ for $y \in [0,1]$. Since the distribution is maximum at $y = 1$, there is a large number of pairs with high similarity that could lie at any point in the character spectra. Hence, the formation of small- and medium-sized clusters in all x regions is expected. This explains the appearance of jumps in the evolution of the largest cluster resulting from the aggregation of small- and medium-sized clusters to G_1. On the other hand, M2 tends to link nodes that lie far from each other in the character spectra. The PDF tells us that the number of optimal pairs for the M2 process (i.e., small or zero S_{ij} and hence y) is low. As a consequence, as the sample size increases, the likelihood that the optimal link already lies within the largest cluster grows, resulting in slow overall growth and a reduction in the formation of medium-sized clusters. Thus, the growth of G_1 for M2 becomes slower and smoother than that of M1.

Figure 1(e) illustrates this fact by examining the size of the largest gap $\Delta G_{\max}$ in the evolution of G_1 [$\Delta G_{\max} = \max\{G_1(L+1) - G_1(L)\}$] due to the addition of a single link. It clearly shows that the gaps associated with M2 are much smaller than those of M1. Interestingly, we find that the gap scales algebraically with the system size as shown in other explosive percolation models [17,18], and this behavior is independent of the link addition mechanism. However, it is found that there is an upper bound in the largest gap for M1 around $1/3$ when the sampling size reaches a specific fraction of the whole system. This observation can be understood as follows. For a uniform character distribution, the mean-field character difference (i.e., $\langle |x_i - x_j| \rangle$) between any two nodes is $1/3$. This means that, for a large sampling size, the optimal pair would carry a character value difference smaller than or around $1/3$. Thus, linking nodes that are separated for more than $1/3$ in character becomes highly unlikely. This process gives rise to the formation of two or three large clusters at separated points of the character spectra that fuse when they have expanded enough, and since the distribution is uniform the growth in size follows the expansion in character and hence the gap is on average $1/3$.

Next, we analyze the diverse aspect, which refers to the character distribution $q(x)$. Toward that end, here we consider $q(x)$ to be equal to a single parameter Beta distribution ($\beta = \alpha$) that is symmetric around $x = 0.5$, as shown in Fig. 2(a). This choice allows us to look at polarized populations but with the same average character value. The uniform distribution previously presented is found for the special case of $\alpha = 1$. Polarized populations can be represented by $0 \leqslant \alpha < 1$, where the severity in the polarization increases as α approaches zero. In the limit of $\alpha = 0$, the system is maximally polarized with half of the nodes having character equal to 1 and half equal to 0. This binary system follows two independent random graph formation processes that will join together only after both graphs are fully formed (see the supplemental material [25]). By contrast, unpolarized populations result for $\alpha > 1$ where the limit of $\alpha \to \infty$ corresponds to the random graph process, which is found since the character values of all nodes are identical.

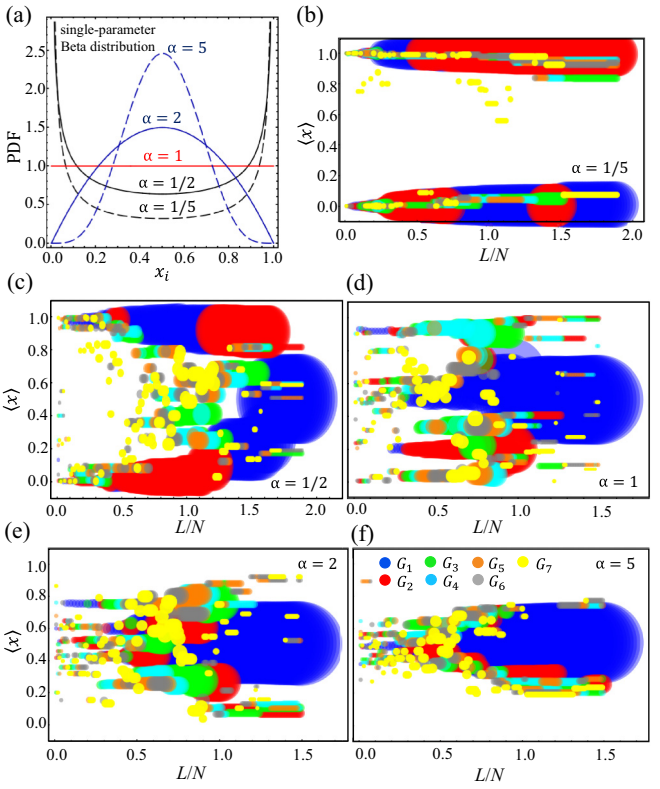

FIG. 2. Effect of diversity in grouping. (a) Probability density function of a Beta distribution for several values of the parameter α for the case when $\alpha = \beta$. (b)–(f) Grouping evolution as a function of the average character of the largest seven clusters (colored bubbles). The size of the bubble is proportional to the square root of the size, and each panel shows a different α value as shown in each panel. The total population is $N = 10^3$ nodes, sampling $k = 10$, and M1 aggregation rule.

Figures 2(b)–2(f) illustrate single simulation results of the evolution of the seven largest clusters (colored bubbles) vertically positioned at their average character value when the population follows a single-parameter Beta distribution with different α values. The size of the bubble is proportional to the square root of the number of nodes within the cluster, and the colors represent the rank from 1 to 7 according to their size [see the legend at Fig. 2(f)]. Here we look at the grouping dynamics for the case of M1, while M2 will be presented in the supplemental material [25]. For $\alpha < 1$, the distribution is highly polarized and groups are formed at opposite extremes of the character spectra. Depending on the severity of the polarization, the system would require more or fewer links to reach consensus around $x = 0.5$. This is because as $\alpha \to 0$, the number of nodes at the center of the x spectra (e.g., moderates) is smaller, and links between groups at opposite poles become less likely [see Fig. 2(b)]. A less severe polarization (e.g., $\alpha = 1/2$) allows for a certain number of moderates, which serve as bridges between the groups in the poles and help reach consensus [see Fig. 2(c)]. Populations with a comparable number of nodes along each portion of the x spectra (i.e., $\alpha = 1$) form small- and medium-sized clusters at different x points that subsequently join together into the largest component around $x = 0.5$. Interestingly, after consensus is attained, some

extreme groups tend to appear at both poles [see Fig. 2(d)]. This behavior is also present for a distribution with a larger amount of moderates than extremists, as depicted in Fig. 2(e) for the case of $\alpha = 2$. Finally, for $\alpha = 5$ we find that the extremes are formed around the edges of the consensus group, as shown in Fig. 2(f). These results can provide some insights concerning how to create consensus (e.g., political) in a diverse and even polarized population of interacting individuals, but they warn of the possibility of leaving residual isolated pockets of individuals with rather extreme average values of x (i.e., away from 0.5).

III. REAL-WORLD NETWORKS

We now explore two different real network systems from the biological and social science domains that experience explosive grouping behavior.

A. Protein homology network

Networks of proteins can be generated by identifying homology relationships, i.e., commonalities in the amino acid sequences of a pair [26,29,30], and connecting them accordingly. Here proteins are viewed as nodes that are linked to each other through weighted edges. In principle, the network can be comprised of all deduced proteins. For simplicity, here we look into the subset of human proteins that have shown features akin to explosive percolation [20]. The link's weight is determined by the homology relationship of a given pair. Highly homologous proteins have a greater weight than heterogeneous proteins. The homology of a given pair and hence the weight of their connecting link is measured by the alignment score s_{ij}, while the score accuracy is determined by the expectation (E) value [31,32]. The smaller the E value, the more reliable the score s_{ij} becomes [31,32]. According to UniProt [33], a total of 159 522 human proteins are deduced, which can be divided into 20 214 that have been reviewed manually against 139 338 that await revision. Here we analyze the scores among the subset of reviewed human proteins provided by the Similarity Matrix of Proteins project (SIMAP) [30], and we have used links with E values up to 10^{-10}. For simplicity, we use the score ratio (SR) as a weight measure, which is defined as $\mathrm{SR} = s_{ij}/\max\{s_{ii}, s_{jj}\}$, where s_{ii} is the self-alignment score. Note that SR is defined within the [0,1] interval, where 1 indicates perfect alignment. The system starts with all proteins isolated, and links are added from high to low according to their alignment. Thus highly homologous communities are formed first, and links between different communities are established later.

Figure 3 presents our findings for the protein system. Figure 3(a) illustrates two snapshots of the evolution of the four largest protein clusters (i.e., G_i for $i = 1,2,3,4$) for a midpoint evolution stage (top panel) where half of the links have been added, and the final stage (bottom panel) where the last link has been added, illustrating their contrasting topologies. Figure 3(b) shows that the size of the largest cluster (red circles) tends to show explosive dynamics as new links are added. The system experiences a roughly linear growth starting near to the midpoint of the dynamics, which accounts for the addition of previously established communities to the largest component. This behavior is captured by our heterogeneous

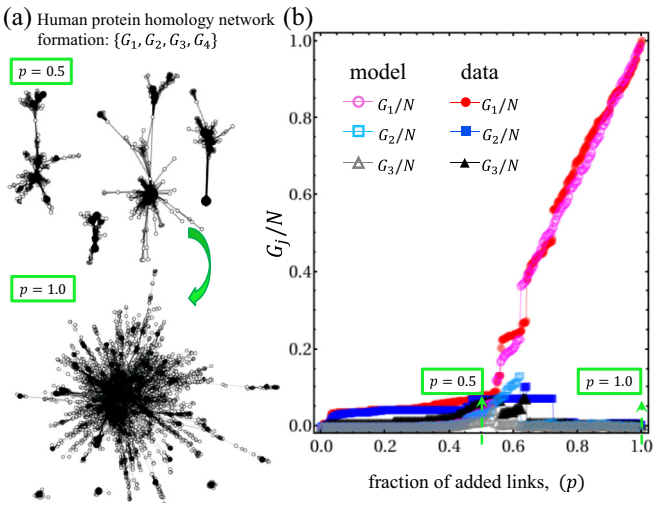

FIG. 3. Heterogeneity in the human protein homology network. (a) Snapshots of the human protein homology network as links are added. The parameter p represents the fraction of added links. The top panel is the early stage for $p = 0.5$ and the bottom panel is the final stage with $p = 1$. Each panel shows the largest four clusters (i.e., G_i for $i = 1,2,3,4$). (b) Evolution of the three largest clusters of the protein network shown in (a) (filled symbols) compared to those of a heterogeneous process (empty symbols) with a uniform character distribution and a sampling size of $k = 8$.

grouping model (pink rings) using a uniform distribution $q(x)$, M1 formation, and a sampling size of $k = 8$. Here we consider only interlink addition since it mimics the addition of the different connected communities that are added to the largest cluster after it has emerged. Both systems experience explosive grouping behavior with comparable rates and gaps. Interestingly, our model also simultaneously reproduces some of the dynamical features of the second and third largest clusters (squares and triangles, respectively). This is largely due to the system's individual heterogeneity, which drives the link addition process according to the affinity (e.g., alignment in the protein system) between the individual nodes. The agreement tends to be higher after the percolation transition than earlier. We attribute this to the strong homology in some protein communities where single nodes can act as hubs gathering many individuals and rising to a non-negligible size. In our model, all nodes are equally likely to be sampled for potential addition. Moreover, the restrictions inflicted by the E value leave many potential links absent, which explains why at the last stage the network is not fully unified [see Fig. 3(a)]. Despite these complications in the alignment parameters, our model is still in reasonable agreement with several features of the evolution of the protein network. This is an indication of the wide flexibility that our model of a heterogeneity framework brings to the grouping behavior, which makes it adaptable to different dynamical systems.

B. Online grouping

We next consider the online social group formation in support of Islamic State (ISIS), whose data were collected in Ref. [34]. This occurs on Europe's largest social media

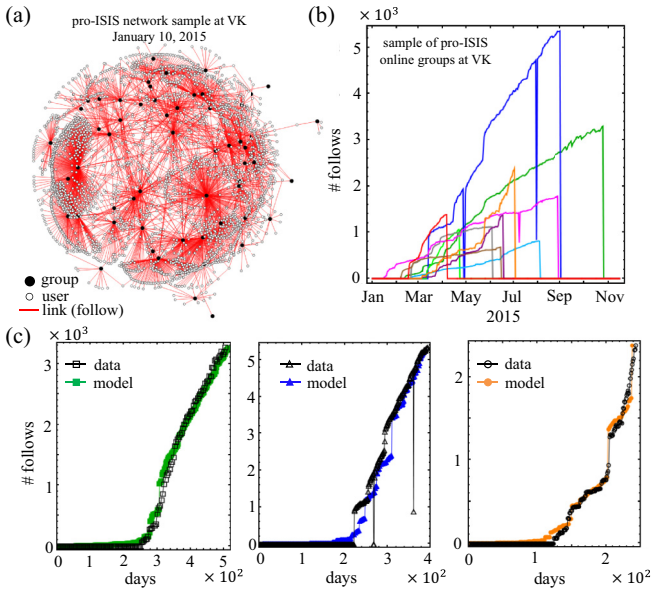

FIG. 4. Explosive grouping in pro-ISIS online groups. (a) Snapshot of online pro-ISIS support groups on the VK platform on January 10, 2015. (b) Evolution of a sample of sharkfin pro-ISIS groups from first detection to shutdown. (c) Examples of explosive behavior in pro-ISIS groups as compared to our heterogeneous model for $k = 5$ (left panel) and $k = 10$ (middle and right panel). For each case, N is set to be the respective group population at the moment of shutdown.

platform based in Russia, VKontakte (VK, https://vk.com). As of December 2017, this site counted 460 million users worldwide, and it has been used by extremists to spread propaganda and to recruit sympathizers. A snapshot of the pro-ISIS network is presented in Fig. 4(a) for January 10, 2015, when 59 different extreme groups were active with a total of 21 881 followers combined, establishing 48 605 connections (i.e., follows). This platform has become ideal for extreme groups, in part because similar networking services such as Facebook shut down these types of online groups almost immediately, while VK takes more time to act. During that active time, the online groups attract followers and grow in particular ways. The methodology of the data extraction is presented in Ref. [34], where different evolutionary adaptations have also been uncovered. For example, groups can change names, restart a new group after a shutdown, or switch between visible and invisible preferences in order to avoid moderators and being shut down, among others.

Figure 4(b) shows the evolution of a sample of extreme online groups from the time of their earliest detection up to the moment when they were shut down, at daily resolution. The size of the groups is determined by the number of users that decide to follow them. We note that in this particular sample the size of the group passes from zero to an average size of 200 follows with a maximum of nearly 1000 follows in a single day. In addition, we see that this irregular growth is repeated at several sections of the formation process. These irregular jumps in size at the start or during the evolution of a given group g_1 could be a consequence of a group (or groups) g_0 being shut down and all its (or their) former members coordinating to either join group g_1, likely due to an affinity with their message, and thus generating a jump in its size, or to open a new group g_2, thus creating a jump at the start of its evolution. Note that the latter process indicates that group g_2 is a continuation of group g_0 and therefore they are essentially the same group, while the former is a cluster aggregation process. In both cases, the changes in size are abrupt, and an association with a random aggregation process is less accurate. We therefore propose that a heterogeneous percolation model with group formation M1 cannot be discarded as a potential mechanism for the creation and subsequent growth of these particular groups.

Figure 4(c) strengthens our proposal by capturing key features of the formation of extreme online groups. The panels show how our heterogeneous model compares with three of the extreme groups shown in Fig. 4(b) (color indicates the specific modeled group). The remaining groups are compared with the model in Fig. S6 of the supplemental material. The model interprets the VK system as a collection of several subsystems, each with a specific subpopulation of potential follows that aggregate over time and whose largest component grows to become the extreme group. Note that the subpopulation is not of users but "follows" since each user can follow several groups simultaneously. Potential followers explore groups daily and decide whether to join or not, arguably based on affinity. This can be considered a competitive process in which only some users (either isolated or from former shut-down groups) add to the extreme group's population. Hence we implement a competitive modeling where on each time step k nodes compete for addition. Our results of Fig. 4(c) as well as Fig. S6 show that values of k between 2 and 10 capture general growth trends as well as some key features, such as the size of the jumps. Due to the bipartite nature of the group evolution, the formation process considers interlink additions only. This framework allows us to estimate the start of the online activity even when the group was invisible or not yet sufficiently extreme, and hence it did not appear in the data collection radar. Also, this modeling paves the way for exploration of different intervention strategies to mitigate the spreading of the group by attacking it, for example at its earliest stage. We note that not all the groups identified can be modeled by this explosive percolation framework for technical reasons, e.g., because of missing data.

IV. SUMMARY

We have shown that a heterogeneous population of interacting individuals can generate explosive grouping behavior. In addition, our model provides a framework to study the impacts of new links on polarized populations. Linking individuals can result in the formation of new residual clusters at the extremes. We also tested our model against two different heterogeneous real-world datasets capturing specific features of the formation process and showing that heterogeneity plays a decisive part in the system's network evolution and should not be ignored.

ACKNOWLEDGMENTS

We thank Chaoming Song for discussions regarding the model, Minzhang Zheng and Yulia Vorobyeva for assistance with the pro-ISIS data, and Thomas Rattei for providing the

protein data. N.F.J. is grateful to the National Science Foundation (NSF) Grant No. CNS1522693 and Air Force (AFOSR) Grant No. FA9550-16-1-0247. The views and conclusions contained herein are solely those of the authors and do not represent official policies or endorsements by any of the entities named in this paper.

[1] M. Anghel, Z. Toroczkai, K. E. Bassler, and G. Korniss, Phys. Rev. Lett. **92**, 058701 (2004).
[2] A. Soulier and T. Halpin-Healy, Phys. Rev. Lett. **90**, 258103 (2003).
[3] B. Goncalves and N. Perra, *Social Phenomena: Data Analytics and Modeling* (Springer, Berlin, 2015).
[4] G. Palla, A. L. Barabasi, and T. Vicsek, Nature (London) **446**, 664 (2007).
[5] E. Estrada, Phys. Rev. E **88**, 042811 (2013).
[6] C. Song, S. Havlin, and H. Makse, Nat. Phys. **2**, 275 (2006).
[7] G. Caldarelli, *Scale-Free Networks: Complex Webs in Nature and Technology* (Oxford University Press, Oxford, 2007).
[8] A. L. Barabasi and H. E. Stanley, *Fractal Concepts in Surface Growth* (Cambridge University Press, Cambridge, 1995).
[9] F. Radicchi and S. Fortunato, Phys. Rev. E **81**, 036110 (2010).
[10] D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (Taylor and Francis, London, 1994).
[11] M. Sahimi, *Applications of Percolation Theory* (Taylor and Francis, London, 1994).
[12] P. Erdös and A. Rényi, Publicationes Mathematicae **6**, 290 (1959).
[13] P. Erdös and A. Rényi, Publications of the Mathematical Institute of the Hungarian Academy of Sciences **5**, 17 (1960).
[14] E. M. Hendriks, M. H. Ernst, and R. M. Ziff, J. Stat. Phys. **31**, 519 (1983).
[15] A. A. Lushnikov, Physica D **222**, 37 (2006).
[16] P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge University Press, Cambridge, 2010).
[17] R. M. D'Souza and J. Nagler, Nat. Phys. **11**, 531 (2014).
[18] J. Nagler, A. Levina, and M. Timme, Nat. Phys. **7**, 265 (2011).
[19] D. Achlioptas, R. M. D'Souza, and J. Spencer, Science **323**, 1453 (2009).
[20] H. D. Rozenfeld, L. K. Gallos, and H. A. Makse, Eur. Phys. J. B **75**, 305 (2010).
[21] G. Bounova and O. de Weck, Phys. Rev. E **85**, 016117 (2012).
[22] T. Nishikawa and A. E. Motter, Phys. Rev. Lett. **117**, 114101 (2016).
[23] Y. Zhang, T. Nishikawa, and A. E. Motter, Phys. Rev. E **95**, 062215 (2017).
[24] Y. Zhang and A. E. Motter, Nonlinearity **31**, R1 (2018).
[25] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.97.032311 for details of the link sampling method, limiting cases, complementary results of the effect of character distribution cluster formation, and additional results for online group evolution.
[26] D. Medini, A. Covacci, and C. Donati, PLoS Comput. Biol. **2**, e173 (2006).
[27] N. F. Johnson, P. Manrique, and P. M. Hui, J. Stat. Phys. **151**, 395 (2013).
[28] P. D. Manrique, P. M. Hui, and N. F. Johnson, Phys. Rev. E **92**, 062803 (2015).
[29] P. F. Jonsson, T. Cavanna, D. Zicha, and P. A. Bates, BMC Bioinform. **7**, 2 (2006).
[30] T. Rattei, R. Arnold, P. Tischler, D. Lindner, V. Stúmpflen, and H. Werner Mewes, Nucl. Acids Res. **34**, D252 (2006).
[31] S. F. Altschul, W. Gish, W. Miller, G. Myers, and D. J. Lipman, J. Mol. Biol. **215**, 403 (1990).
[32] W. R. Pearson, Methods Mol. Biol. **132**, 185 (1999).
[33] R. Apweiler *et al.*, Nucl. Acids Res. **32**, D115 (2004).
[34] N. F. Johnson, M. Zheng, Y. Vorobyeva, A. Gabriel, H. Qi, N. Velasquez, P. Manrique, D. Johnson, E. Restrepo, C. Song, and S. Wuchty, Science **352**, 1459 (2016).