Generalized Gelation Theory Describes Onset of Online Extremist Support

Pedro D. Manrique, Minzhang Zheng, Zhenfeng Cao, Elvira Maria Restrepo, and Neil F. Johnson

Phys. Rev. Lett. **121**, 048301 — Published 27 July 2018

DOI: 10.1103/PhysRevLett.121.048301

Generalized Gelation Theory describes Onset of Online Extremist Support

Pedro D. Manrique[1], Minzhang Zheng[1], Zhenfeng Cao[1], Elvira Maria Restrepo[2], Neil F. Johnson[3]
[1]*Physics Department, University of Miami, Coral Gables, Florida FL 33126, U.S.A.*
[2]*Elliott School of International Affairs, George Washington University, Washington D.C. 20052, U.S.A. and*
[3]*Physics Department, George Washington University, Washington D.C. 20052, U.S.A.*
(Dated: June 27, 2018)

We introduce a generalized form of gelation theory that incorporates individual heterogeneity, and show that it can explain the asynchronous, sudden appearance and growth of online extremist groups supporting ISIS (so-called Islamic State) that emerged globally post-2014. The theory predicts how heterogeneity impacts their onset times and growth profiles, and suggests that online extremist groups present a broad distribution of heterogeneity-dependent aggregation mechanisms centered around homophily. The good agreement between the theory and empirical data, suggests that existing strategies aimed at defeating online extremism under the assumption that it is driven by a few 'bad apples', are misguided. More generally, this generalized theory should apply to a range of real-world systems featuring aggregation among heterogeneous objects.

Aggregation theories developed to date – whether for physical, chemical or biological systems – do not tend to account for the significant heterogeneity found in real-world populations of living objects [1–19]. For example, humans with very different characters have been observed to 'gel' around particular forms of hate speech and extremism remarkably rapidly – for example, the global rise in support of ISIS starting in late 2014 [1, 2]. Indeed, the 'out of the blue' nature of recent attacks in Brussels, Manchester, Paris and London presents security agencies with the fundamental problem of knowing how to move as far as possible left-of-boom in order to detect the *onset* of support for some extremist entity – even if such individuals never end up doing anything in the real world. Even if such aggregates subsequently fragment [20], the dynamics of how they initially emerge and grow, and the consequences of this, are of great interest from both practical and scientific perspectives.

Here we present a generalization of standard aggregation theory [15–19] in order to account for heterogeneity-dependent aggregation dynamics, such as the homophily principle that birds of a feather flock together [21, 22]. We show that it yields good agreement with recent data on online extremism in a way that standard aggregation theory cannot, as well as providing analytic results and insight into the efficacy of individual-based strategies for defeating online extremism. Though we focus on extremism as our empirical testing ground, our results should in principle apply to any system featuring aggregation of heterogeneous objects.

We incorporate the heterogeneity of objects using a variable x assigned to each individual (Fig. 1(a)) [23, 24]. For simplicity we refer to x as a 'character' [23, 24]; we take $0 \leq x \leq 1$; we assign x values by drawing randomly from a distribution $q(x)$ so that each object i has a unique x_i ($i = 1, 2, 3, ..., N$); and we assume x is static over time. All these assumptions can be generalized. Interactions between objects are described in terms of their mutual affinity (i.e. homophily): we define the similarity S_{ij} between i and j as $S_{ij} = 1 - |x_i - x_j|$, so that individuals with alike character have a high similarity

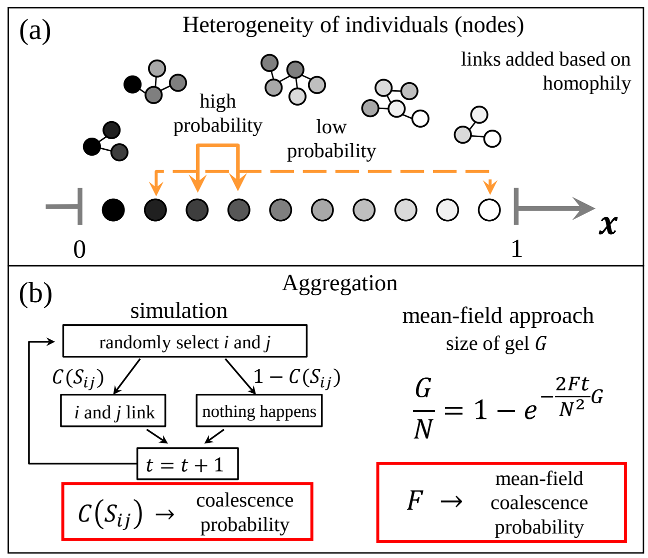

FIG. 1. **Our model of heterogeneity-dependent aggregation** (a) Heterogeneity is modeled by randomly assigning a hidden variable x to each agent, from a given distribution $q(x)$. Link formation between any two individuals depends on their affinity, and is quantified by the coalescence probability $\mathcal{C}(S_{ij})$ which is a function of the similarity $S_{ij} = 1 - |x_i - x_j|$ between any two objects i and j. (b) Flowchart of the aggregation simulation, leading to gel cluster formation (G). Mean-field equation also shown for the gel growth dynamics.

while individuals with unlike character have a low similarity. The aggregation mechanism is quantified by the coalescence probability $\mathcal{C}(S_{ij})$ between individuals i and j (Fig. 1(b)). The limit of random aggregation is obtained by considering all character values to be identical (i.e. $q(x) = \delta(x - x_0)$) representing an entirely homogenous population, or equivalently by making the coalescence probability independent of x (i.e. character-independent): in both cases $\mathcal{C}(S_{ij}) = 1$ and the results of our generalized model reproduce those of traditional gelation theory.

Starting from an isolated population of N objects, clusters form over time by randomly selecting two individuals i and j that merge into a new cluster with a probability $\mathcal{C}(S_{ij})$ or remain separated with a probability of $1-\mathcal{C}(S_{ij})$. Figure 1(b) shows a flow-chart. The process is described analytically at the mean-field level by clusters coalescing at a rate proportional to the product of their sizes (i.e. product kernel) and weighted by a factor F that incorporates the heterogeneity and formation mechanism (see SM for explicit calculation). F determines the likelihood for any pair of elements i and j to merge into a new cluster at a given timestep t for a given population distribution $q(x)$. This generates a set of coupled differential equations for the number of clusters of size s, $n_s(t)$, given by the following for $s \geq 2$ and $s = 1$ respectively:

$$\dot{n}_s(t) = -2F\frac{sn_s}{N^2}\sum_{r=1}^{\infty} rn_r + \frac{F}{N^2}\sum_{r=1}^{s} rn_r(s-r)n_{s-r} \quad (1)$$

$$\dot{n}_s(t) = -2F\frac{n_s}{N^2}\sum_{r=1}^{\infty} rn_r \quad . \quad (2)$$

The first term on the right-hand side of Eqs. 1 and 2 represents the population of clusters of size s that merge with other clusters, while the second term in Eq. 1 is the population of smaller clusters that merge to form clusters of size s. It is known that if this aggregation process is left long enough, a macroscopically observable gel will form [17]. The system undergoes a gelation transition at time $t_c = N/2F$ whose value is determined by a singularity in the second moment of the size distribution (Supplemental Material (SM) [25]).

Our use of a generalized gel-formation framework to describe online extremism, is motivated by the following: It is known [20] that operationally relevant support for pro-ISIS online extremism emerged through macroscopically observable social media groups that appeared suddenly online starting at the end of 2014 and grew out of the 'solute' of several billion online users globally, akin to gel formation [17]. These online groups are each self-contained with each group having its own members and name, and these groups collectively played a key role in terms of building pro-ISIS narratives, recruitment and financing [20, 26, 27]. VKontakte is the largest European social media platform and, like Facebook, has a group tool that enables people with common interests to aggregate together online. While Facebook shuts down such activity, they managed to thrive on VKontakte. Our VKontakte group data collection and datasets are described in full in Ref. 20.

We obtain the temporal evolution of the gel cluster size $G(t)$ by means of the exponential generating function $\mathcal{E}(y,t) \equiv \sum_{s\geq 1} sn_s e^{ys}$ whose partial time derivative takes the form of the inviscid Burgers equation which can be solved by the method of characteristics (see Ref. [17] for the case of homogeneous systems). Above the gel transition point, the formalism predicts that the gel size $G(t)$ obeys

$$G(t) = N\left(1 - e^{-\frac{2Ft}{N^2}G(t)}\right). \quad (3)$$

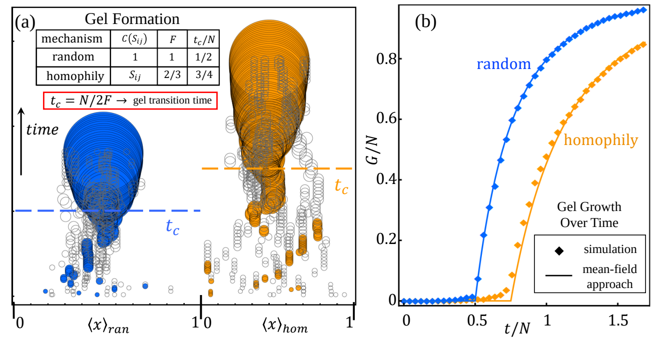

FIG. 2. **Stochastic and analytic mean-field results for the gel dynamics** (a) Table shows analytically derived quantities for our generalized gelation theory based on homophily (subscript *hom*), and for the random version corresponding to traditional gelation theory (i.e. character-independent aggregation, or equivalently a homogenous population with all x values identical, labelled with subscript *ran*). Main panel shows the temporal evolution of clusters for a typical run of the simulation for the two cases (right and left respectively) together with the average x of the growing clusters. Vertical axis is time. Colored disks represent the evolution of gel G while the gray ones are smaller clusters. In all cases the radii grow proportional to $s^{1/2}$. The time limit shown is when G reaches 70% of N ($N = 10^3$ agents). Dashed horizontal lines show the theoretical gel transition times t_c for each case. (b) Results for the simulation (points) and the mean-field formula (lines) for the evolution of G in the two cases.

The solution of Eq. 3 can be expressed in terms of the W-Lambert function as $G/N = 1 - W(z\exp z)/z$ where $z = -2Ft/N$. It can also be shown that the cluster size distribution just before the gelation onset, follows an approximate power-law with exponent $\tau = 5/2$ (see SM):

$$n_s(t \to t_c) \approx \frac{N}{\sqrt{2\pi}} e^{-\frac{s}{2}(1-\frac{t}{t_c})^2} s^{-5/2} \quad (4)$$

where $t_c = N/2F$ contains the F dependence, and hence in turn depends on the group formation mechanism and the character distribution. For a uniform character distribution $q(x)$, the probability density function (PDF) of the similarity $y = S_{ij}$ is $f(y) = 2y$ and hence the mean-field aggregation probability for a process favoring similarity is $F = \int_0^1 yf(y)dy = 2/3$. In the limit of the random model (i.e. homogenous population), $y = 1$ and the character distribution is a Dirac delta which yields $F = 1$. Figure 2 compares the onset and growth of a single gel in our generalized model, to that of the random model corresponding to traditional gelation theory (i.e. character-independent or equivalently an entirely homogenous population). The colored disks in Fig. 2(a) represent the evolution of G while the rings are smaller clusters whose radii are proportional to the square root of their respective size. Figure 2(b) shows the good agreement between the time-evolution of G for each, averaged over 500 realizations (dots), and our mean-field analytic results (solid lines). A fascinating previous study of the

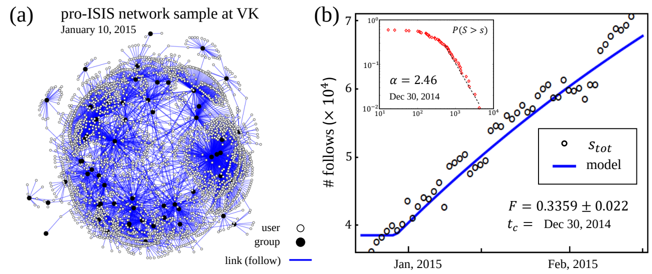

FIG. 3. **Onset dynamics of global pro-ISIS online support** (a) Sample of the online network of groups in support of Islamic State on the VKontakte (VK) platform for an example day: January 10, 2015. Black dots represent groups, white dots are users which are connected through blue links (i.e. follows). (b) Evolution of the total number of follows s_{tot}, i.e. links in the bipartite network shown in (a) (black circles), compared to our analytical model of heterogeneous objects undergoing a gelation process (blue curve). The transition point t_c is found to be December 30, 2014. A fitting of Eq. 3 yields a best-fit value of $F = 0.3349 \pm 0.022$. Moreover for three consecutive days starting December 28, 2014, and hence at $t \to t_c$, we find that the distribution of group sizes follows an approximate power-law distribution with exponent near 5/2 ($\alpha = 2.46$) exactly as predicted by Eq. 4, with a high goodness-of-fit $p \approx 0.64$.

evolution of an online social network is given in Ref. [28], where a dynamical percolation transition was reported in the number of users joining the network and a model was proposed that combines contagion and media influence. Our analysis differs from Ref. [28] in terms of the system studied, and our description of it as one of gelation as opposed to percolation. In our study, the online groups that users choose to join are such that users within the same group are fully connected to each other. Hence, each online group is a self-contained cluster in which all members of the cluster can interact strongly with each other. In Ref. [26], we discuss the contagion process dynamics that emerge at a far later stage in the online development of the groups.

Figure 3(a) shows a snapshot of the pro-ISIS network extracted on January 10th 2015: 59 different social media groups supporting ISIS were found, with a total of 21,881 followers and 48,605 links (i.e. follows). As a result of the extreme content shared in these groups, moderators are constantly chasing them and shutting them down [20, 27, 29–31]. During the period between the end of 2014 and the beginning of 2015, a sudden and roughly continuous growth occurred in the number of added links (i.e. follows) within the whole network, and lasted until mid-2015 where a decay process set in [26]. During the first few weeks of this sudden growth in online extremist support, the number of shutdown events was minimal and hence aggregation processes dominated the system dynamics. This means that this initial period is a good testing ground for our generalized gelation theory.

Figure 3(b) supports our claim that the sudden growth of online pro-ISIS support can be interpreted as a generalized gelation transition from the global online 'solution' of Internet users, and hence can be described by Eqs. 1-4. First, a reasonably well-defined transition point t_c is observed and a fitting of Eq. 3 yields a best-fit value of $F = 0.3349 \pm 0.022$, which is consistent with our generalized model's assumption of values between 0 and 1. Second, as $t \to t_c$ we find from the data that the empirical distribution of group sizes follows an approximate power-law distribution with a high goodness-of-fit ($p \approx 0.64$) and exponent $\alpha = 2.46$ which is very close to the predicted value 5/2 from Eq. 4. Since the system was active for a year prior to t_c, during which multiple group (i.e. cluster) aggregation and fragmentation events could have taken place, it is understandable that a certain level of noise is present within the data. We consider this background noise as the floor from which the gel cluster arises at $t = t_c$, as shown in Fig. 3(b). We scaled the time unit in the model to match that of the data using the gel growth rate during a period of 20 days around the transition point. The fact that the best-fit F for the entire system is close to the value 1/3, could reflect a coalescence process that favors dissimilar individuals (see SM), however more detailed data and content analysis would be required to properly examine this.

We can also apply our generalized gelation theory analysis to examine the formation of each *individual* online group. We focus here on those that are free of any pathological features, i.e. we weed out any groups that temporarily set their public visibility to zero and hence have an apparent group size that jumps temporarily to zero, and we weed out groups that were inactive. We also weed out groups that experience large sudden changes in very short periods of time, since this is more akin to explosive percolation. Since the total number of potential follows varies over time, we restrict the modeling to the first few active weeks where the assumption of a constant subpopulation of follows (i.e. N) holds approximately. We then measured the goodness-of-fit of Eq. 3 against each group's growth. We found a total of 32 groups that give an R-squared higher than 80% during the initial growth period, based on the notion that each group is a gel cluster formed by a subpopulation of follows from a larger pool comprising the entire network. Our results are presented in Fig. 4.

The good agreement shown in Fig. 4 despite the wide range of onset times for when groups first appear, their differing growth rates and their different growth profiles, suggests that our generalized gelation theory is capturing meaningful features of the actual online dynamics. Since our theory is intrinsically a collective many-body one, this suggests that the dynamics of online extremism are collectively driven and hence that proposed solutions aimed at identifying ring-leaders who control and drive the online dynamics, are misguided. This seems good news in that it turns attention to macroscopically observable groups rather than the needle-in-a-haystack problem of identifying a few 'bad' particles in a vast so-

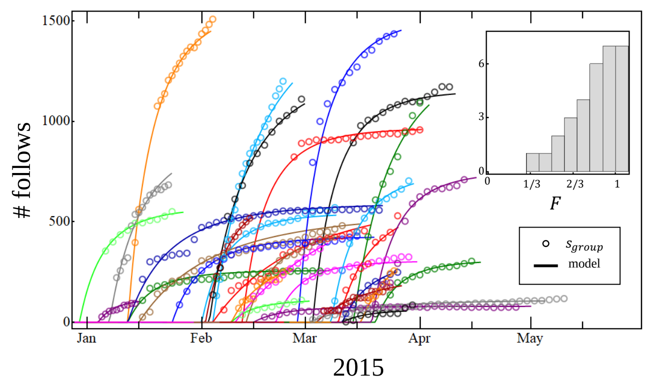

FIG. 4. **Our generalized gelation theory vs. empirical results for onset of online extremist support**. Evolution of the empirically measured group size s_{group} (circles) of individual online pro-ISIS groups, compared to the theoretical curves from our generalized gelation theory (solid lines). The resulting values of the mean-field coalescence probability F for each of the groups, range from 1/3 up to 1, which is consistent with our theory (see inset and text).

lution of several billion online users. For a given group, the onset time and growth profile are characterized by an F value, with these inferred values across all groups shown in Fig. 4 inset. Since existing models of gelation do not include heterogeneity, they all correspond to $F = 1$ and are unable to reproduce the broad spectrum of onset times, growth rates, growth profiles – and in particular the broad distribution of F values – shown in Fig. 4. Without heterogeneity-based aggregation the inset in Fig. 4 would comprise a delta function at $F = 1$. Instead, the actual distribution is centered near $F = 2/3$ in agreement with the idea that the formation of pro-ISIS online extremist groups would be expected to be centered reasonably near the value for homophily (i.e. $F = 2/3$) since homophily is a widely-accepted mechanism for human aggregation. However the distribution's broad nature suggests that online extremist groups present a spectrum of heterogeneity-dependent aggregation mechanisms that are more nuanced versions of homophily. It piles up toward $F = 1$, which is the value where the population is homogeneous (or equivalently, the coalescence mechanism is character-independent). Interestingly, the lower bound occurs near $F = 1/3$ which is the F-value obtained by an aggregation mechanism (c.f. Fig. 1) that favors *dissimilar* individuals (see SM), however there are very few of these. We note that since some groups turn themselves invisible for periods of time, our data necessarily contain some gaps for those particular dates.

Among the limitations of our work is the fact that modeling larger time periods remains a challenge, just as it would in traditional gelation theory, since it involves deriving an expression analogous to Eq. 3 for a population N that changes over time. We have made a first approach to this challenge by considering small linear variations of N in the original equation, i.e. we add a small linear increment to the size ($N(t) = N_0 + kt$) and compare the results to the original case (i.e. $k = 0$). Figure S3 illustrates that this variation resembles the group dynamics for a larger time period than the original modeling. Interestingly there are no significant differences in the estimated value of F, which for the static (i.e. $k = 0$) and dynamical (i.e. $k \neq 0$) versions of the model is $F = 0.97 \pm 0.022$ and $F = 0.96 \pm 0.043$ respectively.

In summary, we have shown that aggregation mechanisms based on individual heterogeneity enrich the dynamics of a finite set of interacting objects, and provide new insight into the urgent societal threat of online extremism. More broadly, this work invites application to the wide range of life science and social systems that involve a heterogeneous population.

We are grateful to Yulia Vorobyeva for detailed help and translation concerning ISIS activity on Russian-language VKontakte, and to Pavel Krapivsky for discussions. NFJ acknowledges funding from the National Science Foundation grant CNS 1522693 and Air Force (AFOSR) grant FA9550-16-1-0247.

[1] J. Silver, P. Horgan, P. Gill. Foreshadowing Targeted Violence: Assessing Leakage of Intent by Public Mass Murderers. *Aggression and Violent Behavior* **38**, 94-100 (2018)

[2] P. Gill. Tactical Innovation and the Provisional Irish Republican Army. *Studies in Conflict and Terrorism* **40**, 573-585 (2017)

[3] D. Medini, A. Covacci, C. Donati. Protein Homology Network Families reveal Step-wise Diversification of Type III and Type IV Secretion Systems. *PLoS Computational Biology* **2** e173, 1543-1551 (2006)

[4] G. Bounova, O. de Weck. Overview of Metrics and their Correlation Patterns for Multiple-Metric Topology Analysis on Heterogeneous Graph Ensembles. *Phys. Rev. E.* **85**, 016117 (2012)

[5] H.R. Pruppacher, J.D. Klett. Microphysics of Clouds and Precipitation (Reidel, Dordrecht, 1978)

[6] S.N. Wall, G.E.A. Anniansson. Numerical Calculations on the Kinetics of Stepwise Micelle Association. *J. Phys. Chem.* **84** 727-736 (1980)

[7] P. V. Coveney, J.A.D. Wattis. Analysis of a Generalized Becker-Döring model of Self-Reproducing Micelles. *Proc. R. Soc. Lond. A.* **452**, 2079-2102 (1996)

[8] J.M. Hidy and J.R. Brock. The Dynamics of Aerocolloidal Systems (Pergamon, New York, 1970)

[9] W. H. Stockmayer. Theory of Molecular Size Distribution and Gel Formation in Branched-Chain Polymers. *J. Chem. Phys.* **11**, 45-55 (1943)

[10] P. Michel, E. Benz, P. Tanga, D.C. Richardson. Collisions and Gravitational Reaccumulation: Forming As-

teroid Families and Satellites. *Science* **294**, 1696-1700 (2001)

[11] P. J. Flory. Principles of Polymer Chemistry (Cornell Univ. Press, Ithaca, 1953)

[12] J.A.D. Wattis. An Introduction to Mathematical Models of Coagulation-Fragmentation Processes: A Discrete Deterministic Mean-Field Approach. *Physica D* **222**, 1-20 (2006)

[13] R.L. Drake. *Topics in Current Aerosol Research.* Vol. 3, G. M. Hidy and J. R. Brock, eds. (Pergamom Press, New York, 1972). Part 2.

[14] M.H. Ernst. *Fractals in Physics.* L. Pietronero and E. Tosatti, eds. (North-Holland, Amsterdam, 1986), p.289

[15] E.M. Hendriks, M.H. Ernst, and R.M. Ziff. Coagulation Equations with Gelation. *J. Stat. Phys.* **31**, 3 (1983)

[16] A.A. Lushnikov. Gelation in Coagulating Systems. *Physica D* **222**, 37-53 (2006)

[17] P. L. Krapivsky, S. Redner and E. Ben-Naim *A Kinetic View of Statistical Physics* (Cambridge University Press, Cambridge, 2010)

[18] R.M. Ziff, E.M. Hendriks and M.H. Ernst. Critical Properties for Gelation: A Kinetic Approach. *Phys. Rev. Lett.* **49**, 8 (1982)

[19] E.M. Hendriks, M.H. Ernst and R.M. Ziff. Coagulation Equations with Gelation. *J. Stat. Phys.* **31**, 3 (1983)

[20] N.F. Johnson, M. Zheng, Y. Vorobyeva, A. Gabriel, N. Velasquez, P. Manrique, D. Johnson, E. Restrepo, C. Song and S. Wuchty. New Online Ecology of Adversarial Aggregates: ISIS and Beyond. *Science* **352**, 1459-1463 (2016)

[21] P. F. Lazarsfeld, R. K. Merton. Friendship as a Social Process: a Substantive and Methodological Analysis, in *Freedom and Control in Modern Society* (Berger, New York, 1954). pp. 18-66

[22] M. McPherson, L. Smith-Lovin, J. M. Cook. Birds of a Feather: Homophily in Social Networks. *Annu. Rev. Sociol.* **27**, 415-44 (2001)

[23] N.F. Johnson, P. Manrique and P.M. Hui. Modeling Insurgent Dynamics Including Heterogeneity: A Statistical Physics Approach. *J. Stat. Phys.* **151**, 395 (2013)

[24] P.D. Manrique, P.M. Hui and N.F. Johnson. Internal Character Dictates Phase Transition Dynamics between Isolation and Cohesive Grouping. *Phys. Rev. E* **92**, 062803 (2015)

[25] See Supplemental Material (SM) for fuller details including derivations. The SM also includes Refs. [32-41].

[26] M. Zheng, Z. Cao, Y. Vorobyeva, P. Manrique, C. Song, N.F. Johnson. Multiscale Dynamical Network Mechanisms Underlying Aging from Birth to Death. *Scientific Reports* **8**, 3552 (2018)

[27] BBC News. Russian Students Targeted as Recruits by Islamic State. Available at `http://www.bbc.com/news/world-europe-33634214` (July 24, 2015)

[28] K. Kleineberg, M. Boguñá. Evolution of the Digital Society Reveals Balance between Viral and Mass Media Influence. *Phys. Rev. X.* **4**, 031046 (2014)

[29] BBC News. Islamic State Web Accounts to be Blocked by New Police Team. Available at `http://www.bbc.com/news/world-europe-33220037` (June 22, 2015)

[30] J.M. Berger and H. Perez. The Islamic State's Diminishing Returns on Twitter. GW Program on Extremism (2016). Available at `https://cchs.gwu.edu/sites/cchs.gwu.edu/files/downloads/Berger_Occasional%20Paper.pdf`

[31] J. Paraszczuk. Why Are Russian, Central Asian Militants Vanishing From Social Networks? RadioFreeEurope November 05, 2015. Available at `http://www.rferl.org/a/russian-central-asian-militants-vanish-socialnetworks/27347535.html`

[32] P.G.J. van Dongen and M.H. Ernst. *J. Stat. Phys.* **49**, 889-926 (1987)

[33] P.J. Flory. Molecular Size Distribution in Three Dimensional Polymers. I. Gelation. *J. Am. Chem. Soc.* **63**, 3083 (1941)

[34] W.H. Stockmayer. Theory of Molecular Size Distribution and Gel Formation in Branched Polymers II. General Cross Linking. *Journal of Chemical Physics.* **12**, 125 (1944)

[35] D. Stauffer, A. Aharony. *Introduction to Percolation Theory* (Taylor & Francis, New York, 1994)

[36] M. Sahimi. *Applications of Percolation Theory* (Taylor & Francis, New York, 1994)

[37] P. Erdös, A. Rényi. On Random Graphs I. Math. *Debrecen* **6**, 290-297 (1959)

[38] P. Erdös, A. Rényi. On the Evolution of Random Graphs *Publ. Math. Inst. Hungar. Acad. Sci.* **5**, 17-61 (1960)

[39] J.C. Bohorquez, S. Gourley, A. Dixon, M. Spagat and N.F. Johnson. Common ecology quantifies human insurgency. *Nature* **462**, 911 (2009)

[40] A. Clauset, M. Young, K.S. Gleditsch. On the Frequency of Severe Terrorist Events. J. Confl. Resolut. **51**, 5887 (2007)

[41] A. Clauset, K. Gleditsch. The Developmental Dynamics of Terrorist Organizations. *PLoS One* **7**, e48633 (2012)