Corrected 7 January 2019. See full text.

SOCIAL NETWORKS

The strength of long-range ties in population-scale social networks

Patrick S. Park[1]*, Joshua E. Blumenstock[2], Michael W. Macy[3,4]*

Long-range connections that span large social networks are widely assumed to be weak, composed of sporadic and emotionally distant relationships. However, researchers historically have lacked the population-scale network data needed to verify the predicted weakness. Using data from 11 culturally diverse population-scale networks on four continents—encompassing 56 million Twitter users and 58 million mobile phone subscribers—we find that long-range ties are nearly as strong as social ties embedded within a small circle of friends. These high-bandwidth connections have important implications for diffusion and social integration.

Over the last 40 years, the social sciences have embraced the counterintuitive thesis that individuals are more likely to acquire new information from a weak social tie to an acquaintance than from a strong tie with a close friend or family member (*1*). The reason is straightforward: Information that one acquires from within a "small circle of friends" is more likely to be redundant than information acquired from an acquaintance in a distant region of a social network. Thus, the prevailing consensus, dating back to Granovetter's seminal thesis (*1*), is that there is a trade-off between the diversity of information acquired through weak bridging ties (linking individuals whose social circles do not overlap) and the volume of information, or bandwidth, acquired through strong, structurally embedded ties (between individuals with at least one friend in common) (*2*–*4*).

This diversity-bandwidth trade-off is empirically supported by studies showing that tie strength decreases as social ties become less embedded, i.e., as they connect individuals with fewer network "neighbors" in common (*5*). By extrapolation, tie strength should decrease further between individuals who do not even have "neighbors of neighbors" in common; that is, as the network distance, or range of the tie, increases. Tie range is defined as the second-shortest path length, i.e., the number of intermediary ties required to reach from an individual node to its neighbor if their direct tie were removed (*1*, *6*) (Fig. 1).

The diversity-bandwidth trade-off has been widely tested and confirmed for small networks with embedded and unembedded short-range ties (*2*). Until recently, however, it has been difficult to empirically test whether the trade-off applies to long-range ties because of the inability to obtain data for the population-scale social networks in which long-range ties can be found. In theory, a long-range tie could be found in a relatively small ring lattice. However, the tendency for social networks to be highly clustered means that long range-ties are rarely observed in networks with no more than a few thousand nodes, such as villages, schools, and workplaces. Thus, the existence of long-range ties has been largely a postulate of the "small world" puzzle of "six degrees of separation," which refers to the minimum number of intermediate ties between any two people on the planet. Watts and Strogatz (*7*) used simulations to show that the six degrees phenomenon could be explained by long-range ties, but the ties themselves were never directly observed. Since then, several studies have confirmed the six (or fewer) degrees of separation in a variety of contexts, including email networks (*8*), MSN Messenger (*9*), and Facebook (*10*). Although these results are consistent with the postulated existence of long-range ties, their prevalence and strength have not been directly measured, and other studies demonstrate that heavy-tailed degree distributions could also account for six degrees, even in the absence of long-range ties (*11*, *12*).

We report direct evidence of long-range ties in social networks, made possible by analyzing 11 population-scale communication networks from culturally and economically diverse populations spanning four continents: three independent nationwide phone networks (in Afghanistan, Rwanda, and a large European country), as well as 56 million Twitter users in eight countries (fig. S1) with relatively high Twitter penetration (United States, United Kingdom, France, Netherlands, Japan, South Korea, Singapore, and Turkey). Details of data collection and measurement are provided in supplementary materials (SM) section 1.

The data confirmed, at a global scale, previous findings that social ties tend to be weaker (lower call volume and fewer tweet exchanges) when people share fewer common neighbors (*1*, *5*, *13*). This was evident in all 11 networks (fig. S2). However, our focus was on testing whether tie strength declined with tie range. Tie strength declined as range increased from two (the theoretical lower bound) to four (the upper bound on what is likely to be observable in small local networks) in all three phone networks and most of the Twitter networks (Fig. 2).

What happened above range four, the region that is difficult to observe without population-scale networks, was especially notable. Instead of declining further, tie strength increased with the network distance spanned, especially in the phone networks (Fig. 2B). Figure S3 shows that ties with range six or greater were approximately as strong as embedded ties with one common neighbor in all three phone networks and in three out of the eight Twitter networks (Japan, South Korea, and the Netherlands).

We refer to these high-bandwidth long-range ties as network "wormholes," borrowing the term from cosmology to capture the possibility that, though relatively rare (fig. S4), long-range ties can provide high-bandwidth shortcuts across vast reaches of network space. To illustrate, Fig. 3 depicts Singapore's Twitter network (the smallest of the networks), in which a tie is composed of one or more reciprocated @mentions. The wormholes, defined here as ties above range six and above median tie strength, are shown with curved yellow edges and represent only 0.46% of all ties. The inset shows how network wormholes can substantially shrink networks by directly

[1]Ross School of Business, University of Michigan, Ann Arbor, MI 48109, USA. [2]School of Information, University of California, Berkeley, Berkeley, CA 94720, USA. [3]Information Science Department, Cornell University, Ithaca, NY 14853, USA. [4]Department of Sociology, Cornell University, Ithaca, NY 14853, USA.
*Corresponding author. Email: patpark@umich.edu (P.P.); mwm14@cornell.edu (M.M.)

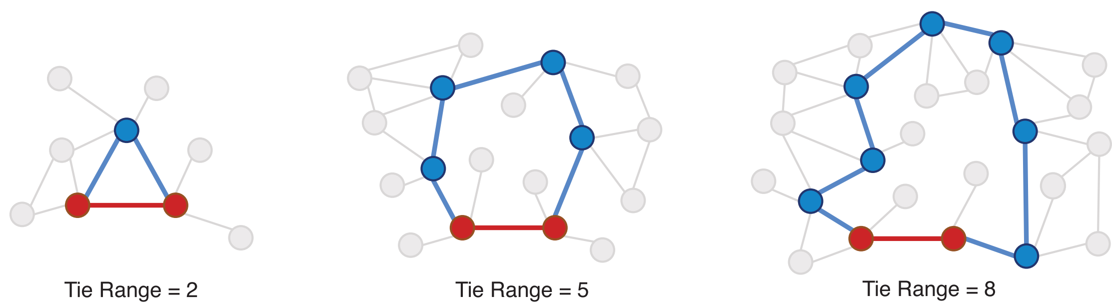

Fig. 1. Tie range is defined as the second-shortest (blue) path length between two connected (red) nodes.

Corrected 7 January 2019. See full text.

linking nodes that would otherwise be connected by a long chain of intermediary links.

The unexpected strength of long-range ties has implications for two related puzzles on the diffusion of new information. Figure 4 addresses these puzzles by decomposing Fig. 2 into the within- and between-individual variation. The within-individual analysis compares the strength of each individual's short- and long-range ties. This analysis addresses Granovetter's (*1*) original puzzle: From whom are we more likely to receive new information? By contrast, the between-individual analysis compares the average tie strength between people who have mostly long-range ties with those whose ties are mostly short-range. This corresponds more closely to the follow-up puzzle posed by Aral and Van Alstyne (*2*): Who is more likely to receive new information? The U-shape pattern was more pronounced in the within-individual analysis than the overall results in Fig. 2; across all 11 networks, people interacted with their most socially distant neighbors nearly as much as they did with their embedded neighbors. The between-individual U-shape was less consistent across networks (see SM section 2).

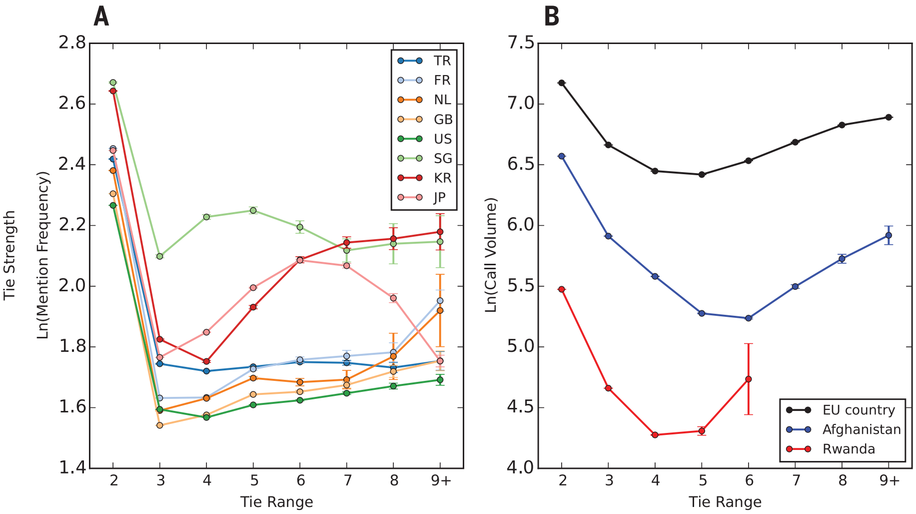

Fig. 2. The strength of social ties by tie range. Results are shown for eight Twitter networks (**A**) and three phone networks (**B**). Tie strength [mean and 99% confidence interval (CI)] is measured as the log of the frequency of bidirected @mentions (A) and the log of total bidirected call volume in seconds (B).

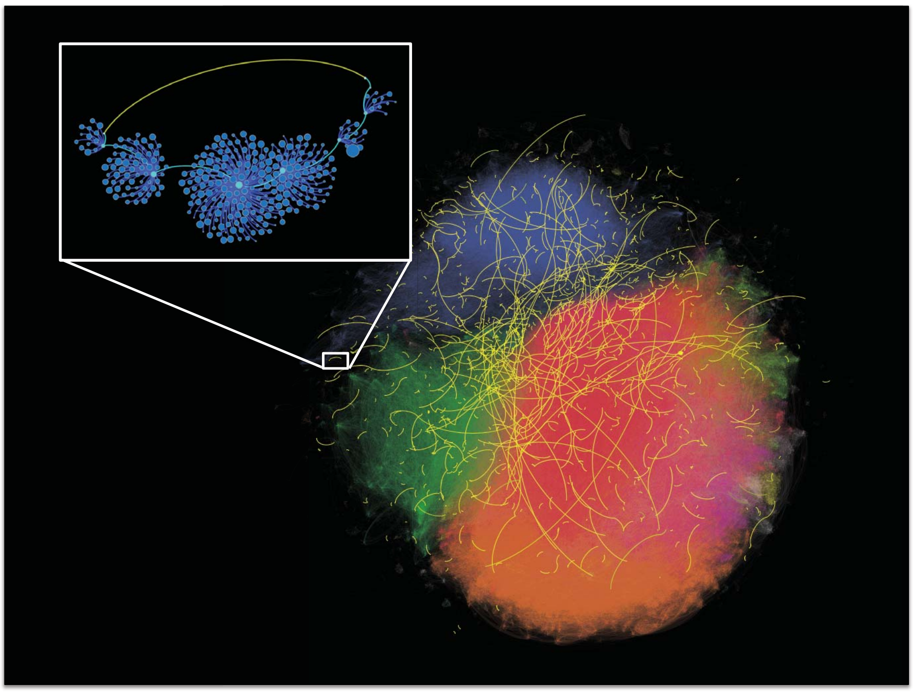

Fig. 3. Network wormholes in Singapore's Twitter network. Each dot represents an individual, and each edge represents a bidirected @mention. Nodes and edges are colored according to membership in distinct network communities (*31*). A sample of network wormholes (with range six or above and above-median tie strength) is shown in yellow. The inset highlights a single wormhole of range eight, i.e., the second-shortest path between the yellow nodes requires traversing eight intermediary ties (blue edges). The sizes of the nodes in the inset are proportional to the number of network neighbors.

To verify the robustness of these findings, we ran a battery of tests, described in detail in SM section 3. To begin, Fig. 2 reports the generality of our results across alternative communication platforms and across countries with widely divergent cultural and economic conditions. Twitter and phone networks differ fundamentally in user demographics, relational structure (multilateral distribution versus dyadic conversation), mode of expression (text versus voice), population penetration (partial versus full), openness (public versus private), and incremental cost (free versus paid). Despite these differences, we nevertheless observed the same phenomenon: strong ties that span extreme network distances. This ubiquity suggests that the result does not reflect an idiosyncrasy of the country or communication platform, such as the opportunity on Twitter to form strong relationships with erstwhile strangers, a preference for unembedded relationships in individualistic cultures, or demographic biases in technology use.

We also tested robustness across several alternative measures of tie strength (SM sections 3.2 to 3.4): the mean duration and frequency of calls on the phone networks (fig. S5), the affective strength of message content (fig. S6), and the reciprocity of @mentions (fig. S7) on Twitter (*1, 14*). In all instances, we observed that tie strength eventually increased with range, confirming the pattern in Fig. 2. Finally, we found little support for the possibility that the results were an artifact of missing data (see SM section 3.5). In principle, strong embedded ties could be incorrectly measured as network wormholes if data were missing on common neighbors. This possibility is mitigated by the existence of network wormholes in all 11 observed networks despite differences in population coverage, from approximately 3.5% of the 2014 French internet population on Twitter, to more than 90% of all phone lines in the European phone network. Nevertheless, we tested the effects of missing data by randomly removing nodes and edges from the observed networks. We found that missing data do not explain the strength of long-range ties (fig. S8) or cause embedded ties to appear to be long-range (figs. S9 and S10). The reason is straightforward: For a range two tie to appear to be range three because of missing data, all common neighbors would need to be missing;

Corrected 7 January 2019. See full text.

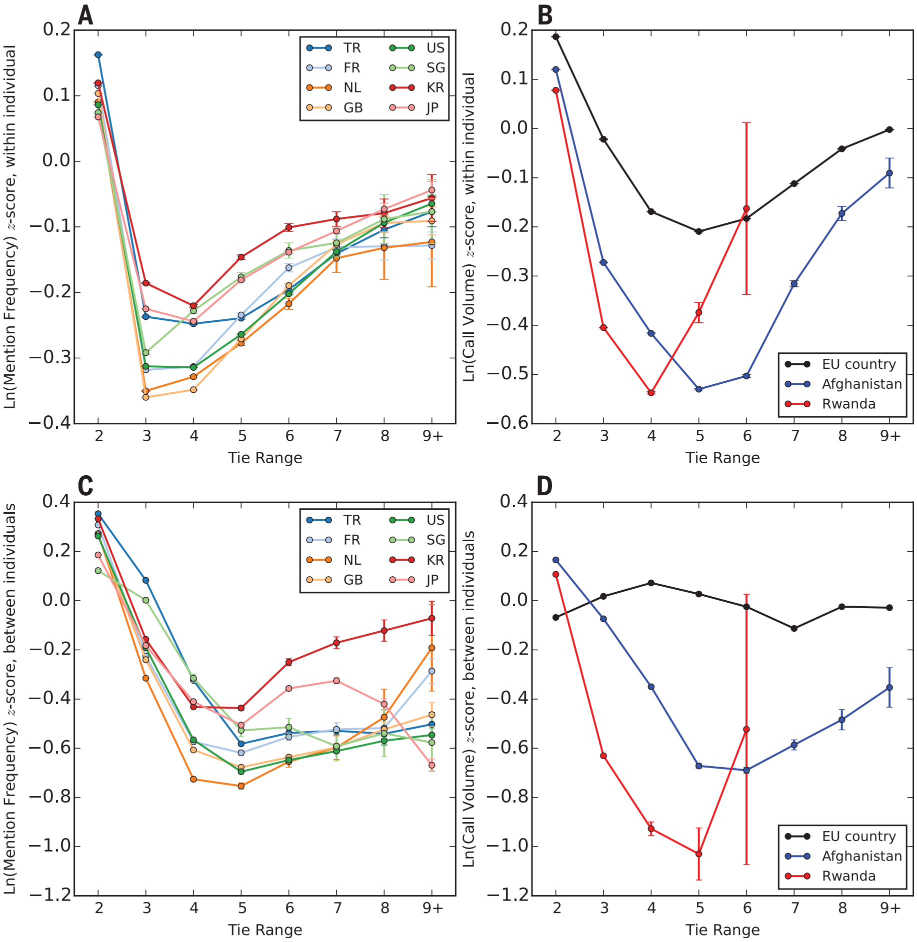

Fig. 4. Within-individual and between-individual decomposition of tie strength (mean and 99% CI) by tie range. The first row shows the within-individual relationship for Twitter (**A**) and phone networks (**B**), where the z-score is calculated by standardizing each tie with the individual's average and standard deviation of tie strength. The second row shows the between-individual relationship for Twitter (**C**) and phone networks (**D**), where the z-score is calculated by standardizing each individual's average tie strength with the grand mean and standard deviation of the entire network. The tie range in the second row represents the average range of each individual's ties, rounded to the nearest integer.

if even a single common neighbor were observed, the tie would remain range two. For the range to appear longer than three, the number of necessary missing neighbors or missing ties increases exponentially.

The discovery of high-bandwidth ties that span vast network distances poses intriguing puzzles that call for further research into their formation and surprising strength. Although a comprehensive investigation is beyond the scope of this report, SM section 4 explores the cultural context, spatial distance, social function, and personal attributes of these ties to look for possible clues, which we briefly summarize.

First, the content of the messages exchanged over strong, long-range Twitter ties displayed no single characteristic pattern; see table S1 for a few example conversations. Topic modeling of message content suggested that network wormholes frequently involved religious and cultural topics (tables S2 to S4) as well as social process words (e.g., "buddy," "talk"), but very few work-related words (e.g., "job", "boss"). See SM section 4 and fig. S11 for details.

Second, temporal analysis suggested that network wormholes were more likely to be interpersonal social relationships rather than instrumental or work-related (e.g., between a service provider and client). In particular, the increase in tie strength with longer range was driven by ties that were active during non–working hours (fig. S12).

Finally, the strength of long-range ties was not a byproduct of physical distance (SM section 4.3). Prior work has shown that tie probability declines with geographic distance (*15*–*18*), which may have helped promote the widely held but historically untestable assumption that tie strength decreases with range. Figures S13 and S14 show that physical and network distances were conceptually and empirically distinct dimensions. Results were consistent with previous findings that tie strength generally decreases with spatial distance, but the pattern was the opposite for network distance. Notably, the change in tie strength with range largely followed the patterns in Fig. 2, even among ties with shorter spatial distance.

Future research should target three possible explanations for the formation and strength of long-range ties. First, long-range ties frequently connected low-degree nodes on the periphery of the network (fig. S15). This may indicate that limited time or attention induced people to choose between a small number of close friends and many weakly tied acquaintances (*19*), and those with few neighbors had fewer chances to have neighbors (or neighbors of neighbors) in common. Second, Burt (*20*) found that weak ties are more likely to break over time. If social and spatial mobility breaks weaker ties, the stronger ones that remain become longer range as weaker indirect paths erode (see SM section 4.4). For example, in a book about his friendship with his high school calculus teacher, Strogatz (*21*) tells the story of their strong tie that remained strong despite the increasing separation of their evolving social networks over time. This winnowing process might also explain the heavy-tailed range distribution (fig. S16). Finally, research on multiplexity and multidimensional homophily (*22*) indicates that social networks tend to be composed of many different types of relationships (friendship, kinship, work, politics, religion, hobbies, etc.). The discovery of network wormholes suggests that these layers may not be fully integrated, e.g., a strongly tied religious or political neighbor might not be introduced to one's workplace colleagues (*23*).

The surprising strength of long-range ties was found in a wide range of cultures, communication platforms, and measures of network structure and survived a battery of robustness tests. But do these network wormholes matter, given their relative rarity? SM section 5 presents a counterfactual experiment that compares the observed Singapore network with an otherwise identical network in which tie strengths were permuted inversely with range (as would be expected with a diversity-bandwidth trade-off). The counterfactual network greatly increased the average shortest path length (i.e., the mean strength-weighted geodesic distance) between two random nodes, relative to the observed network with wormholes (fig. S17). In simulation experiments, contagions also spread more slowly and reached fewer nodes when wormholes were removed from the network (fig. S18). These effects, combined with the tendency for network wormholes to link peripheral nodes, support recent studies that question the dependence of diffusion on "hubs" (*24*, *25*). Finally, the stronger emotional affect observed in longer-range Twitter ties highlights the potential implications for the spread of emotional contagions (*13*, *26*) such as moral indignation, political celebration, ideological fervor, happiness, and value judgments (*27*) that in turn may influence voting (*28*), participation in risky social movements, and health (*29*, *30*).

REFERENCES AND NOTES

1. M. S. Granovetter, *Am. J. Sociol.* **78**, 1360–1380 (1973).
2. S. Aral, M. Van Alstyne, *Am. J. Sociol.* **117**, 90–171 (2011).
3. R. Reagans, B. McEvily, *Adm. Sci. Q.* **48**, 240–267 (2003).
4. A. Tiwana, *Strateg. Manage. J.* **29**, 251–272 (2008).
5. J.-P. Onnela *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7332–7336 (2007).
6. G. Kossinets, D. J. Watts, *Science* **311**, 88–90 (2006).
7. D. J. Watts, S. H. Strogatz, *Nature* **393**, 440–442 (1998).
8. P. S. Dodds, R. Muhamad, D. J. Watts, *Science* **301**, 827–829 (2003).
9. J. Leskovec, E. Horvitz, in *Proceeding of the 17th international conference on World Wide Web - WWW '08* (ACM Press, New York, 2008), pp. 915–924.
10. L. Backstrom, P. Boldi, M. Rosa, J. Ugander, S. Vigna, in *Proceedings of the 3rd Annual ACM Web Science Conference on - WebSci '12* (ACM Press, New York, 2012), pp. 33–42.
11. A.-L. Barabási, R. Albert, *Science* **286**, 509–512 (1999).
12. L. A. Amaral, A. Scala, M. Barthelemy, H. E. Stanley, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11149–11152 (2000).
13. A. D. I. Kramer, J. E. Guillory, J. T. Hancock, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8788–8790 (2014).
14. E. Gilbert, K. Karahalios, in *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09* (ACM Press, New York, 2009), pp. 211–220.
15. G. Daraganova *et al.*, *Soc. Networks* **34**, 6–17 (2012).
16. R. Lambiotte *et al.*, *Physica A* **387**, 5317–5325 (2008).
17. D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, A. Tomkins, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 11623–11628 (2005).
18. P. Preciado, T. A. B. Snijders, W. J. Burk, H. Stattin, M. Kerr, *Soc. Networks* **34**, 18–31 (2012).
19. J. Saramäki *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 942–947 (2014).
20. R. S. Burt, *Soc. Networks* **22**, 1–28 (2000).
21. S. Strogatz, *The Calculus of Friendship* (Princeton Univ. Press, 2009).
22. M. McPherson, L. Smith-Lovin, J. Cook, *Annu. Rev. Sociol.* **27**, 415–444 (2001).
23. P. Block, T. Grund, *Netw. Sci. (Camb Univ Press)* **2**, 189–212 (2014).
24. S. González-Bailón, *Decoding the Social World: Data Science and the Unintended Consequences of Communication* (MIT Press, 2017).
25. P. Barberá *et al.*, *PLOS ONE* **10**, e0143611 (2015).
26. J. H. Fowler, N. A. Christakis, *BMJ* **337**, a2338 (2008).
27. L. Muchnik, S. Aral, S. J. Taylor, *Science* **341**, 647–651 (2013).
28. R. M. Bond *et al.*, *Nature* **489**, 295–298 (2012).
29. D. McAdam, *Am. J. Sociol.* **92**, 64–90 (1986).
30. D. Centola, *Science* **329**, 1194–1197 (2010).
31. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008).
32. P. S. Park, J. E. Blumenstock, M. W. Macy, Replication Data for: "The Strength of Long-Range Ties in Population-Scale Social Networks," Harvard Dataverse, Version 1 (2018); https://doi.org/10.7910/DVN/NPRNCC

ACKNOWLEDGMENTS

We are especially grateful to N. Eagle, who provided access to phone data for this study. We thank I. Kloumann, B. Lee, and M. Park for helpful feedback and N. Bhat, M. Flashman, X. Ma, S. Ogden, and Y. Zhao for research assistance. **Funding:** This study was supported by the U.S. National Science Foundation (SES-1226483 and SES-1434164), the Air Force Office of Scientific Research (FA9550-15-1-0036), the Defense Advanced Research Project Agency (HR0011-18-C0049), and the National Research Foundation of Korea (NRF-2016S1A3A2925033). **Author contributions:** M.W.M. and P.S.P. designed the study; J.E.B., M.W.M., and P.S.P. collected the data; P.S.P. analyzed the data; and J.E.B., M.W.M., and P.S.P. wrote the manuscript. **Competing interests:** All authors declare no competing interests. **Data and materials availability:** The de-identified edgelists with edge weight, tie range, and geographic distance to construct the eight Twitter networks and related scripts are available at (*32*).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6421/1410/suppl/DC1
Materials and Methods
Figs. S1 to S18
Tables S1 to S4
References (*33*–*55*)

2 August 2018; accepted 6 November 2018
10.1126/science.aau9735