



Community Structure in Time-Dependent, Multiscale, and Multiplex Networks

Peter J. Mucha *et al.*

Science **328**, 876 (2010);

DOI: 10.1126/science.1184819

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of November 2, 2013):

A correction has been published for this article at:

<http://www.sciencemag.org/content/329/5989/277.3.full.html>

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/328/5980/876.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2010/05/13/328.5980.876.DC1.html>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/content/328/5980/876.full.html#related>

This article **cites 19 articles**, 3 of which can be accessed free:

<http://www.sciencemag.org/content/328/5980/876.full.html#ref-list-1>

This article has been **cited by** 13 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/content/328/5980/876.full.html#related-urls>

This article appears in the following **subject collections**:

Computers, Mathematics

http://www.sciencemag.org/cgi/collection/comp_math

Community Structure in Time-Dependent, Multiscale, and Multiplex Networks

Peter J. Mucha,^{1,2*} Thomas Richardson,^{1,3} Kevin Macon,¹ Mason A. Porter,^{4,5} Jukka-Pekka Onnela^{6,7}

Network science is an interdisciplinary endeavor, with methods and applications drawn from across the natural, social, and information sciences. A prominent problem in network science is the algorithmic detection of tightly connected groups of nodes known as communities. We developed a generalized framework of network quality functions that allowed us to study the community structure of arbitrary multislice networks, which are combinations of individual networks coupled through links that connect each node in one network slice to itself in other slices. This framework allows studies of community structure in a general setting encompassing networks that evolve over time, have multiple types of links (multiplexity), and have multiple scales.

The study of graphs, or networks, has a long tradition in fields such as sociology and mathematics, and it is now ubiquitous in academic and everyday settings. An important tool in network analysis is the detection of mesoscopic structures known as communities (or cohesive groups), which are defined intuitively as groups of nodes that are more tightly connected to each other than they are to the rest of the network (1–3). One way to quantify communities is by a quality function that compares the number of intracommunity edges to what one would expect at random. Given the network adjacency matrix A , where the element A_{ij} details a direct connection between nodes i and j , one can construct a quality function Q (4, 5) for the partitioning of nodes into communities as $Q = \sum_{ij} (A_{ij} - P_{ij}) \delta(g_i, g_j)$, where $\delta(g_i, g_j) = 1$ if the community assignments g_i and g_j of nodes i and j are the same and 0 otherwise, and P_{ij} is the expected weight of the edge between i and j under a specified null model.

The choice of null model is a crucial consideration in studying network community structure (2). After selecting a null model appropriate to the network and application at hand, one can use a variety of computational heuristics to assign nodes to communities to optimize the quality Q (2, 3). However, such null models have not been available for time-dependent networks; analyses have instead depended on ad hoc methods to

piece together the structures obtained at different times (6–9) or have abandoned quality functions in favor of such alternatives as the Minimum Description Length principle (10). Although tensor decompositions (11) have been used to cluster network data with different types of connections, no quality-function method has been developed for such multiplex networks.

We developed a methodology to remove these limits, generalizing the determination of community structure via quality functions to multislice networks that are defined by coupling multiple adjacency matrices (Fig. 1). The connections encoded by the network slices are flexible; they can represent variations across time, variations across different types of connections, or even community detection of the same network at different scales. However, the usual procedure for establishing a quality function as a direct count of the intracommunity edge weight minus that

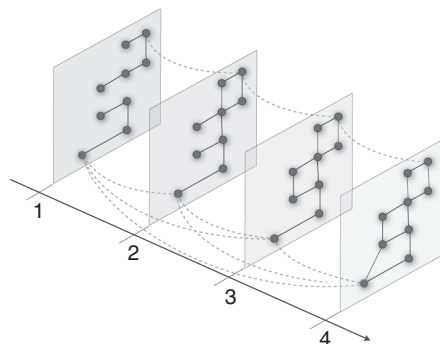


Fig. 1. Schematic of a multislice network. Four slices $s = \{1, 2, 3, 4\}$ represented by adjacencies A_{ij} encode intraslice connections (solid lines). Interslice connections (dashed lines) are encoded by C_{jrs} , specifying the coupling of node j to itself between slices r and s . For clarity, interslice couplings are shown for only two nodes and depict two different types of couplings: (i) coupling between neighboring slices, appropriate for ordered slices; and (ii) all-to-all interslice coupling, appropriate for categorical slices.

expected at random fails to provide any contribution from these interslice couplings. Because they are specified by common identifications of nodes across slices, interslice couplings are either present or absent by definition, so when they do fall inside communities, their contribution in the count of intra-community edges exactly cancels that expected at random. In contrast, by formulating a null model in terms of stability of communities under Laplacian dynamics, we have derived a principled generalization of community detection to multislice networks,

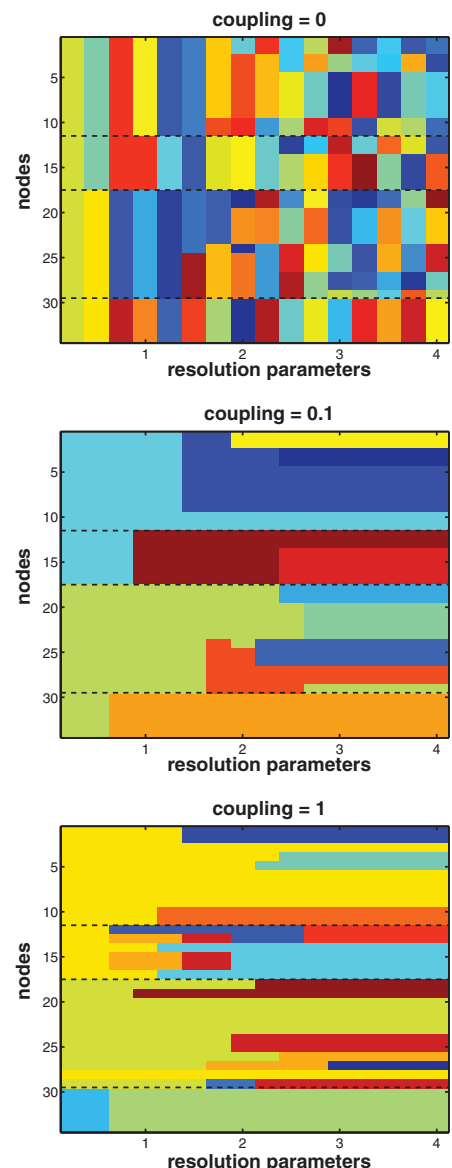


Fig. 2. Multislice community detection of the Zachary Karate Club network (22) across multiple resolutions. Colors depict community assignments of the 34 nodes (renumbered vertically to group similarly assigned nodes) in each of the 16 slices (with resolution parameters $\gamma_s = \{0.25, 0.5, \dots, 4\}$), for $\omega = 0$ (top), $\omega = 0.1$ (middle), and $\omega = 1$ (bottom). Dashed lines bound the communities obtained using the default resolution ($\gamma = 1$).

¹Carolina Center for Interdisciplinary Applied Mathematics, Department of Mathematics, University of North Carolina, Chapel Hill, NC 27599, USA. ²Institute for Advanced Materials, Nanoscience and Technology, University of North Carolina, Chapel Hill, NC 27599, USA. ³Operations Research, North Carolina State University, Raleigh, NC 27695, USA. ⁴Oxford Centre for Industrial and Applied Mathematics, Mathematical Institute, University of Oxford, Oxford OX1 3LB, UK. ⁵CABDyN Complexity Centre, University of Oxford, Oxford OX1 1HP, UK. ⁶Department of Health Care Policy, Harvard Medical School, Boston, MA 02115, USA. ⁷Harvard Kennedy School, Harvard University, Cambridge, MA 02138, USA.

*To whom correspondence should be addressed. E-mail: mucha@unc.edu

with a single parameter controlling the interslice correspondence of communities.

Important to our method is the equivalence between the modularity quality function (12) [with a resolution parameter (5)] and stability of communities under Laplacian dynamics (13), which we have generalized to recover the null models for bipartite, directed, and signed networks (14). First, we obtained the resolution-parameter generaliza-

tion of Barber's null model for bipartite networks (15) by requiring the independent joint probability contribution to stability in (13) to be conditional on the type of connection necessary to step between two nodes. Second, we recovered the standard null model for directed networks (16, 17) (again with a resolution parameter) by generalizing the Laplacian dynamics to include motion along different kinds of connections—in this case,

both with and against the direction of a link. By this generalization, we similarly recovered a null model for signed networks (18). Third, we interpreted the stability under Laplacian dynamics flexibly to permit different spreading weights on the different types of links, giving multiple resolution parameters to recover a general null model for signed networks (19).

We applied these generalizations to derive null models for multislice networks that extend the existing quality-function methodology, including an additional parameter ω to control the coupling between slices. Representing each network slice s by adjacencies A_{ijs} between nodes i and j , with interslice couplings C_{jrs} that connect node j in slice r to itself in slice s (Fig. 1), we have restricted our attention to unipartite, undirected network slices ($A_{ijs} = A_{jis}$) and couplings ($C_{jrs} = C_{jsr}$), but we can incorporate additional structure in the slices and couplings in the same manner as demonstrated for single-slice null models. Notating the strengths of each node individually in each slice by $k_{js} = \sum_i A_{ijs}$ and across slices by $c_{js} = \sum_r C_{jrs}$, we define the multislice strength by $\kappa_{js} = k_{js} + c_{js}$. The continuous-time Laplacian dynamics given by

$$\dot{p}_{is} = \sum_{jr} \frac{(A_{ijs} \delta_{sr} + \delta_{ij} C_{jsr}) p_{jr}}{\kappa_{jr}} - p_{is} \quad (1)$$

respects the intraslice nature of A_{ijs} and the interslice couplings of C_{jsr} . Using the steady-state probability distribution $p_{jr}^* = \kappa_{jr}/2\mu$, where $2\mu = \sum_{jr} \kappa_{jr}$, we obtained the multislice null model in terms of the probability $\rho_{is|jr}$ of sampling node i in slice s conditional on whether the multislice structure allows one to step from (j, r) to (i, s) , accounting for intra- and interslice steps separately as

$$\rho_{is|jr} p_{jr}^* = \left(\frac{k_{is}}{2m_s} \frac{k_{jr}}{\kappa_{jr}} \delta_{sr} + \frac{C_{jsr}}{c_{jr}} \frac{C_{jrs}}{\kappa_{jr}} \delta_{ij} \right) \frac{\kappa_{jr}}{2\mu} \quad (2)$$

where $m_s = \sum_{jr} \kappa_{js}$. The second term in parentheses, which describes the conditional probability of motion between two slices, leverages the definition of the C_{jsr} coupling. That is, the conditional probability of stepping from (j, r) to (i, s) along an interslice coupling is nonzero if and only if $i = j$, and it is proportional to the probability C_{jsr}/κ_{jr} of selecting the precise interslice link that connects to slice s . Subtracting this conditional joint probability from the linear (in time) approximation of the exponential describing the Laplacian dynamics, we obtained a multislice generalization of modularity (14):

$$\mathcal{Q}_{\text{multislice}} = \frac{1}{2\mu} \sum_{ijsr} \left[\left(A_{ijs} - \gamma_s \frac{k_{is} k_{js}}{2m_s} \right) \delta_{sr} + \delta_{ij} C_{jsr} \right] \delta(g_{is}, g_{jr}) \quad (3)$$

where we have used reweighting of the conditional probabilities, which allows a different resolution γ_s in each slice. We have absorbed the resolution parameter for the interslice couplings into the magnitude of the elements of C_{jsr} , which, for simplicity, we presume to take binary values $\{0, \omega\}$ indicating the absence (0) or presence (ω) of interslice links.

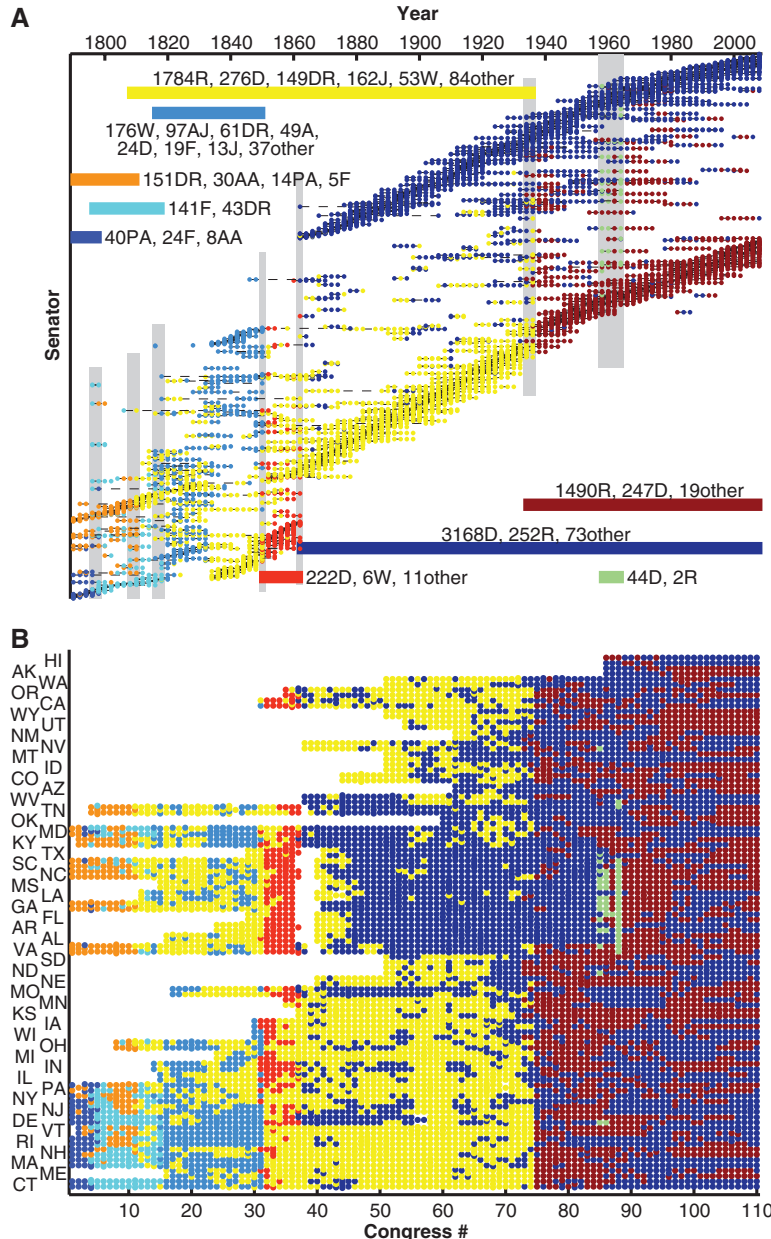


Fig. 3. Multislice community detection of U.S. Senate roll call vote similarities (23) with $\omega = 0.5$ coupling of 110 slices (i.e., the number of 2-year Congresses from 1789 to 2008) across time. **(A)** Colors indicate assignments to nine communities of the 1884 unique senators (sorted vertically and connected across Congresses by dashed lines) in each Congress in which they appear. The dark blue and red communities correspond closely to the modern Democratic and Republican parties, respectively. Horizontal bars indicate the historical period of each community, with accompanying text enumerating nominal party affiliations of the single-slice nodes (each representing a senator in a Congress): PA, pro-administration; AA, anti-administration; F, Federalist; DR, Democratic-Republican; W, Whig; AJ, anti-Jackson; A, Adams; J, Jackson; D, Democratic; R, Republican. Vertical gray bars indicate Congresses in which three communities appeared simultaneously. **(B)** The same assignments according to state affiliations.

Table 1. Communities in the first wave of the multiplex “Tastes, Ties, and Time” network (24), using the default resolution ($\gamma = 1$) in each of the four slices of data (Facebook friendships, picture friendships, roommates, and housing groups) under various couplings ω across slices, which changed the number of communities and percentages of individuals assigned on a per-slice basis to one, two, three, or four communities.

ω	Number of communities	Communities per individual (%)			
		1	2	3	4
0	1036	0	0	0	100
0.1	122	14.0	40.5	37.3	8.2
0.2	66	19.9	49.1	25.3	5.7
0.3	49	26.2	48.3	21.6	3.9
0.4	36	31.8	47.0	18.4	2.8
0.5	31	39.3	42.4	16.8	1.5
1	16	100	0	0	0

Community detection in multislice networks can then proceed using many of the same computational heuristics that are currently available for single-slice networks [although, as with the standard definition of modularity, one must be cautious about the resolution of communities (20) and the likelihood of complex quality landscapes that necessitate caution in interpreting results on real networks (21)]. We studied examples that have multiple resolutions [Zachary Karate Club (22)], vary over time [voting similarities in the U.S. Senate (23)], or are multiplex [the “Tastes, Ties, and Time” cohort of university students (24)]. We provide additional details for each example in (14).

We performed simultaneous community detection across multiple resolutions (scales) in the well-known Zachary Karate Club network, which encodes the friendships between 34 members of a 1970s university karate club (22). Keeping the same unweighted adjacency matrix across slices ($A_{ijs} = A_{ij}$ for all s), the resolution associated with each slice is dictated by a specified sequence of γ_s parameters, which we chose to be the 16 values $\gamma_s = \{0.25, 0.5, 0.75, \dots, 4\}$. In Fig. 2, we depict the community assignments obtained for coupling strengths $\omega = \{0, 0.1, 1\}$ between each neighboring pair of the 16 ordered slices. These results simultaneously probe all scales, including the partition of the Karate Club into four communities at the default resolution of modularity (3, 25). Additionally, we identified nodes that have an especially strong tendency to break off from larger communities (e.g., nodes 24 to 29 in Fig. 2).

We also considered roll call voting in the U.S. Senate across time, from the 1st Congress to the 110th, covering the years 1789 to 2008 and including 1884 distinct senator IDs (26). We defined weighted connections between each pair of senators by a similarity between their voting, specified independently for each 2-year Congress (23). We studied the multislice collection of these 110 networks, with each individual senator coupled to himself or herself when appearing in consecutive Congresses. Multislice community detection uncovered interesting details about the continuity of individual and group voting trends over time that are not captured by the union of the 110 independent partitions of the separate Congresses.

Figure 3 depicts a partition into nine communities that we obtained using coupling $\omega = 0.5$. The Congresses in which three communities appeared simultaneously are each historically noteworthy: The 4th and 5th Congresses were the first with political parties; the 10th and 11th Congresses occurred during the political drama of former Vice President Aaron Burr’s indictment for treason; the 14th and 15th Congresses witnessed the beginning of changing group structures in the Democratic-Republican party amidst the dying Federalist party (23); the 31st Congress included the Compromise of 1850; the 37th Congress occurred during the beginning of the American Civil War; the 73rd and 74th Congresses followed the landslide 1932 election (during the Great Depression); and the 85th to 88th Congresses brought the major American civil rights acts, including the congressional fights over the Civil Rights Acts of 1957, 1960, and 1964.

Finally, we applied multislice community detection to a multiplex network of 1640 college students at a northeastern American university (24), including symmetrized connections from the first wave of this data representing (i) Facebook friendships, (ii) picture friendships, (iii) roommates, and (iv) student housing-group preferences. Because the different connection types are categorical, the natural interslice couplings connect an individual in a slice to himself or herself in each of the other three network slices. This coupling between categorical slices thus differs from that above, which connected only neighboring (ordered) slices. Table 1 indicates the numbers of communities and the percentages of individuals assigned to one, two, three, or four communities across the four types of connections for different values of ω , as a first investigation of the relative redundancy across the connection types.

Our multislice framework makes it possible to study community structure in a much broader class of networks than was previously possible. Instead of detecting communities in one static network at a time, our formulation generalizing the Laplacian dynamics approach of (13) permits the simultaneous quality-function study of community structure across multiple times, multiple resolution parameter values, and multiple types of links. We

used this method to demonstrate insights in real-world networks that would have been difficult or impossible to obtain without the simultaneous consideration of multiple network slices. Although our examples included only one kind of variation at a time, our framework applies equally well to networks that have multiple such features (e.g., time-dependent multiplex networks). We expect multislice community detection to become a powerful tool for studying such systems.

References and Notes

1. M. Girvan, M. E. J. Newman, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7821 (2002).
2. M. A. Porter, J.-P. Onnela, P. J. Mucha, *Not. Am. Math. Soc.* **56**, 1082 (2009).
3. S. Fortunato, *Phys. Rep.* **486**, 75 (2010).
4. M. E. J. Newman, *Phys. Rev. E* **74**, 036104 (2006).
5. J. Reichardt, S. Bornholdt, *Phys. Rev. E* **74**, 016110 (2006).
6. J. Hopcroft, O. Khan, B. Kulis, B. Selman, *Proc. Natl. Acad. Sci. U.S.A.* **101** (suppl. 1), 5249 (2004).
7. T. Y. Berger-Wolf, J. Saia, in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2006), p. 523 (10.1145/1150402.1150462).
8. G. Palla, A.-L. Barabási, T. Vicsek, *Nature* **446**, 664 (2007).
9. D. J. Fenn *et al.*, *Chaos* **19**, 033119 (2009).
10. J. Sun, C. Faloutsos, S. Papadimitriou, P. S. Yu, in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2007), p. 687 (10.1145/1281192.1281266).
11. T. M. Selee, T. G. Kolda, W. P. Kegelmeyer, J. D. Griffin, *CSRI Summer Proceedings 2007, Technical Report SAND2007-7977*, Sandia National Laboratories, Albuquerque, NM and Livermore, CA, M. L. Parks, S. S. Collis, Eds. (2007), p. 87 (www.cs.sandia.gov/CSRI/Proceedings/).
12. M. E. J. Newman, M. Girvan, *Phys. Rev. E* **69**, 026113 (2004).
13. R. Lambiotte, J. C. Delvenne, M. Barahona, <http://arxiv.org/abs/0812.1770> (2008).
14. See supporting material on Science Online.
15. M. J. Barber, *Phys. Rev. E* **76**, 066102 (2007).
16. A. Arenas, J. Duch, A. Fernandez, S. Gomez, *N. J. Phys.* **9**, 176 (2007).
17. E. A. Leicht, M. E. J. Newman, *Phys. Rev. Lett.* **100**, 118703 (2008).
18. S. Gómez, P. Jensen, A. Arenas, *Phys. Rev. E* **80**, 016114 (2009).
19. V. A. Traag, J. Bruggeman, *Phys. Rev. E* **80**, 036115 (2009).
20. S. Fortunato, M. Barthélemy *Proc. Natl. Acad. Sci. U.S.A.* **104**, 36 (2007).
21. B. H. Good, Y.-A. de Montjoye, A. Clauset, *Phys. Rev. E* **81**, 046106 (2010).
22. W. W. Zachary, *J. Anthropol. Res.* **33**, 452 (1977).
23. A. S. Waugh, L. Pei, J. H. Fowler, P. J. Mucha, M. A. Porter, <http://arxiv.org/abs/0907.3509> (2009).
24. K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, N. Christakis, *Soc. Networks* **30**, 330 (2008).
25. T. Richardson, P. J. Mucha, M. A. Porter, *Phys. Rev. E* **80**, 036111 (2009).
26. K. T. Poole, Voteview (<http://voteview.com>) (2008).
27. We thank N. A. Christakis, L. Meneades, and K. Lewis for access to and helping with the “Tastes, Ties, and Time” data; S. Reid and A. L. Traud for help developing code; and A. Clauset, J.-C. Delvenne, S. Fortunato, M. Gould, and V. Traag for discussions. Congressional roll call data are from <http://voteview.com> (26). Supported by NSF grant DMS-0645369 (P.J.M.), James S. McDonnell Foundation grant 220020177 (M.A.P.), and the Fulbright Program (J.-P.O.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5980/876/DC1
SOM Text
References
17 November 2009; accepted 22 March 2010
10.1126/science.1184819

ERRATUM

Post date 16 July 2010

Reports: “Community structure in time-dependent, multiscale, and multiplex networks” by P. J. Mucha *et al.* (14 May, p. 876). Equation 3 contained a typographical error that was not caught during the editing process: The δ_{sr} term should have been outside of the parentheses within the square brackets. The correct equation, which also appears in the supporting online material as equation 9, is as follows:

$$Q_{\text{multislice}} = \frac{1}{2\mu} \sum_{ijsr} \left[\left(A_{ijs} - \gamma_s \frac{k_{is}k_{js}}{2m_s} \right) \delta_{sr} + \delta_{ij} C_{jsr} \right] \delta(g_{is}, g_{jr})$$

See the revised supporting online material (www.sciencemag.org/cgi/content/full/sci;328/5980/876/DC2), which also includes a correction to equation 11. The computations supporting the examples described in the Report were all performed with the correct formula for $Q_{\text{multislice}}$. The authors thank Giuseppe Mangioni for pointing out the error.