Identifying Labor Market Areas Based on Link Communities

Stephan J. Goetz
Professor
Northeast Regional Center for Rural Development; National Agricultural and Rural Development Policy Center, Pennsylvania State University
Department of Agricultural Economics and Rural Sociology, Pennsylvania State University
207-C Armsby, The Pennsylvania State University, University Park, PA 16802-5602, USA
sgoetz@psu.edu

Yicheol Han
Corresponding author
Post-Doctoral fellow
Northeast Regional Center for Rural Development; National Agricultural and Rural Development Policy Center, Pennsylvania State University
7G Armsby, The Pennsylvania State University, University Park, PA 16802-5602, USA
Email: yuh14@psu.edu

Selected Paper prepared for presentation at the 2015 Agricultural & Applied Economics Association and Western Agricultural Economics Association Annual Meeting, San Francisco, CA, July 26-28

Identifying Labor Market Areas Based on Link Communities

Abstract: Labor Market Areas (LMAs) are distinctive communities of counties within a nation's commuting network that are closely connected with one another than with other counties. The overlapping of communities within a hierarchical structure is a crucial feature of real-world networks, including commuting, yet existing methods are inadequate for modeling such overlapping because the concept is inconsistent with hierarchical ordering. Ahn *et al*. (2010) introduced the link community method to address this problem but did not consider weighted and directed links such as commuting flows. In this paper, we extend the link community method to accommodate directed and weighted networks using the idea that edges can be presented as vectors. We then apply our proposed method to U.S. commuting data. Results suggest that our new method reliably identifies the LMAs that we would expect to find. In our case, however, these LMAs can also overlap one another.

Keywords: labor market area, LMA, link community, complex networks, commuting

1. Introduction

Labor market areas (LMAs) are integrated spatial units within which residents may change jobs without having to change their residence. Alternatively, they are "integrated sets of cities and their surrounding suburban hinterlands across which labor and capital can be reallocated at very low cost" (Florida *et al*., 2008 p.459). Counties within an LMA share common economic activities and there is greater homogeneity of counties within than across LMAs, and the

counties are exposed to similar economic growth and development forces. A key analytical challenge in this context remains that of identifying and defining LMAs with consistency and rigor.

Researchers and federal agencies in the U.S. have used commuting data to delineate labor market areas. For example, Tolbert and Sizer (1996) identified commuting zones and labor market areas from 1990 U.S. county-to-county commuting flow data (see also Goetz, 2014). Because commuting is closely associated with population and workforce agglomerations, commuting networks provide a plausible theoretical foundation for categorizing LMAs. Yet to date, the idea of using network analysis to identify LMAs has not been explored. Goetz *et al*. (2010) use the fact that commuting networks involve nodes (counties) and links (highways and residence to place-of-work commuter flows across counties) to study how network spillovers and a county's position in the network affect economic growth, but they do not identify LMAs.

Within network science, robust methods have recently been developed to detect communities or clusters of nodes that have certain features in common, such as greater frequency of contact or above-average traffic flow volumes. A key challenge with these approaches is that they do not allow individual nodes (e.g., counties) to belong to multiple groups, communities or, in our case, labor market areas. In other words, these groups are mutually exclusive and overlapping membership is ruled out. Yet a given county may belong to multiple commuting sheds. For example, a commuter county on the East Coast that sends workers to Washington, DC, Philadelphia, and Baltimore may be located in the border region between the cores of multiple LMAs, and it is not obvious into which of these commuting regions the county should be classified. Alternatively, it is implausible that New York City belongs to only a single LMA, as is the case with the current classification method.

Recently Ahn *et al*. (2010) proposed focusing on the common characteristics of links to cluster components of a network instead of examining the nodes. They show that communities or groups defined using links rather than nodes can reveal multi-scale complexity in networks that range from biochemical to linguistic and social. They write, (p.761) "the fact that many real networks have communities with pervasive overlap, where each and every node belongs to more than one group, has the consequence that a global hierarchy of nodes cannot capture the relationships between overlapping groups." To circumvent this problem, Ahn *et al.* (2010) propose the concept of a "link community," which allows for overlapping nodes within a hierarchy.[1] The very nature of a network implies interactions between nodes, and the link community actually contains more information than does a purely node-based community, and it potentially provides new insights into different networks types.

Our contributions in this paper are two-fold. First, we apply the idea of communities of links (as opposed to nodes) to the problem of identifying LMAs using commuting network data. Second, we develop a method of identifying communities with common features (such as belonging to the same LMA) when the underlying network data consist of weighted and directed links. This represents a key innovation beyond the work of Ahn *et al*. (2010). We employ a link vector (defined below) to represent commuting flows, using the dot product between two vectors and the *cosine* function to compare angles between commuting vectors. To our knowledge, this relationship has not previously been used within network science, and with certainty not in this context.

[1] Thus, while link communities identified using hierarchical clustering methods do not allow for overlapping links, they do accommodate overlapping nodes.

2. Literature Review and Motivation

The complexity of systems derives from the inter-knitted nature of interactions among network members. Networks (and systems) typically contain components within which nodes are relatively more similar or more densely connected to each other than to other nodes in the network (Watts and Strogatz, 1998; Wasserman and Faust, 1994, p. 249). These components are usually referred to as communities, subgroups, clusters, cohesive groups or modules, and various methods have been developed to detect such components, including motifs (Milo *et al.*, 2002), cliques (Wasserman and Faust, 1994, p.254), edge removal (Girvan and Newman, 2002), and probabilistic clustering (Bacher, 2000). At the heart of most clustering or grouping algorithms is the idea that within-group ties among nodes should be stronger and the grouped node are more "similar" in some respect.

The hierarchical clustering method (HCM), based on dendrograms, is often used to identify clusters because it accommodates different community shapes (e.g., line, ring, star and tree, etc. networks) without requiring prior knowledge of the number or size of clusters within the network. When larger communities are recursively built on smaller ones, the underlying networks take on a hierarchical structure (Ravasz *et al.*, 2002; Ravasz and Barabási, 2003; Newman 2004). Indeed, Tolbert and Sizer (1996) apply the HCM to county-to-county commuting data to identify labor market areas (communities) based on the sum of shared commuters between counties and the number of resident labor forces in the counties. The clean analytical split into distinct communities fails, however, when a node can belong to multiple groups, which produces overlapping communities (Palla *et al.*, 2005 and 2007). For example, the HCM divides a network into sub-groups whereby every node belongs to only one community, within a

hierarchical structure.[2] Yet the overlapping of communities within a hierarchical network structure is a crucial feature of many real-world networks (Lancichinetti *et al*., 2009; Shen *et al.*, 2009).

To circumvent this fundamental contradiction between overlap and hierarchy, recent studies have proposed the concept of a "link community." This approach focuses on the common characteristics of edges rather than the nodes to group or cluster elements of a network into communities that are at once overlapping and exhibit a hierarchy among nodes (Evans and Lambiotte, 2009 and 2010; Ahn *et al*., 2010). Importantly, while link communities identified using the HCM cannot accommodate the overlapping of edges, they do allow the overlapping of nodes.

3. Method

The link community method introduced by Ahn *et al*. (2010) treats any two nodes connected by a link as a set, or one object. Here an n by n node matrix (e.g., the 3,143 by 3,143 matrix of county-to-county commuting flows) is transformed into an $n(n\text{-}1)$ by $n(n\text{-}1)$ link matrix. Any one row (or column) can then be compare to all other rows (columns) and the results can be ranked in terms of similarity. Our contribution in extending the method of Ahn *et al*. (2010) is to measure the degree of similarity between two links using the degree to which they are parallel in vector space, as we explain below. Our proposed approach uses information on the direction (e.g., county A sends commuters to county B but not vice versa) or weight (the number of commuters) of a link, and it can accommodate weights of negative values (e.g., reverse

[2] Note that, in contrast, such overlapping is possible in non-hierarchical clusters.

commuting). Therefore, our approach can be applied to a wider range of real-world networks than is possible with existing methods.

To define our link communities, we need to introduce the tool of link vectors in the next section. This link vector allows us to efficiently characterize a link and the two nodes it connects. Second, we classify two link vectors as "similar" using the characteristics of their edges. Third, we use a partition density and dendrogram to determine when a link vector "belongs" to a community and when it does not. As noted our procedure hierarchically sorts links into mutually exclusive groups while allowing the nodes to be members of overlapping communities.

3.1 Link vector

A node is defined, among other factors, by its connections with neighboring nodes within a network. Because these connections can be characterized in matrix form, as noted above, the branches of a node -- i.e., the rows (or columns) of a matrix -- can be represented as a vector. In the same way, a link can be defined by the edges attached to its end nodes, as follows. We designate keystone k as a midpoint node between two link vectors: $\boldsymbol{e_{ik}}$ and $\boldsymbol{e_{jk}}$ (Fig. 1), which are described by the connected edges as in Eq. (1):

$$\boldsymbol{e_{ik}} = w_{is}(1-\delta_{ik})(1-\delta_{ij})\forall s \quad , \quad \boldsymbol{e_{jk}} = w_{js}(1-\delta_{jk})(1-\delta_{ij})\,\forall s \tag{1}$$

where w_{is} is the weight of the connection from node i to s, and δ_{ij} the Kronecker delta (δ_{ij}=1 when $i = j$, and zero otherwise). The three nodes i, j, and k are distinct; note that we need at least three nodes to create one link community. The minimum weight of flow is zero, which

means that two nodes are not connected. To minimize the influence of long distance commuting and to ensure robust connections between counties, we include link vector $\boldsymbol{e_{ik}}$ only when its weight w_{ik} exceeds a minimum threshold.[3]

3.2 Similarity

Next, we need to define when two link communities are similar; this allows us to classify them into more homogenous communities. Vectors have both direction and magnitude (weight) and two vectors are similar when they are parallel to each other. We use the angle between them to define a similarity index. By definition, dot product $\boldsymbol{e_{ik}} \cdot \boldsymbol{e_{jk}}$ of two vectors $\boldsymbol{e_{ik}}$ and $\boldsymbol{e_{jk}}$ is the product of their magnitudes and the cosine of the angle θ between them measured at the point where the two vectors originate, as shown in Eq. (2), which we have rearranged and solved for $cos(\theta)$:

$$s_{ij,k} = \frac{\boldsymbol{e_{ik}} \cdot \boldsymbol{e_{jk}}}{|\boldsymbol{e_{ik}}||\boldsymbol{e_{jk}}|} = \frac{\sum_s e_{ik}(s) \cdot e_{jk}(s)}{\sqrt{\sum_s e_{ik}(s)^2}\sqrt{\sum_s (e_{jk}(s))^2}} = cos\,\theta \quad (2)$$

Here, $s_{ij,k}$ is the similarity between $\boldsymbol{e_{ik}}$ and $\boldsymbol{e_{jk}}$, and k is the keystone node that two link vector shares. The $e_{ik}(s)$ is the s^{th} element of link vector $\boldsymbol{e_{ik}}$. The quantity $cos\theta$ can be used an index that measures the similarity of two edges in a network. Similarity ranges from 1 (very similar) when the two link vectors are parallel to 0 (very dissimilar) when the link vectors are orthogonal. An example of the similarity index is shown in Fig. 2.

[3] We use 10 as the minimum threshold. If the threshold changes, the communities detected would also change.

3.3 Partition density

To detect communities that consist of similar edges, we start with a bottom-up, agglomerative hierarchical clustering algorithm. Initially, each edge stands alone and a single edge cannot comprise a community. As the criterion for what is considered similar declines or becomes less strict from the maximum (i.e., 1), only the most highly similar edge pairs are at first merged into communities. As the similarity requirement continues to fall, we recursively merge edges (or subgroups of edges) in descending order of similarity until all edges would eventually belong to a single community. In this hierarchical clustering process, we still need to determine when an optimal link community structure has been reached (Fig. 3). In other words, we want to choose a threshold that provides a classification of edges so that those belonging to a community are sufficiently similar, and yet we do not want an unmanageably large number of communities (or in our case, LMAs).

The threshold at which to cut the continuous similarity index can be determined using *partition density D*. This partition density measures the cohesion between edges in communities based on how "clique-ish" vs. "tree-ish" each link in the community is (Ahn *et al.*, 2010 and Fig. 4). A well-organized community has a densely connected structure and a high partition density *D*, and vice versa. Denoting the number of edges in community c as m_c and the number of nodes as n_c, the *link density* D_c of community c is calculated from Eq. (3).

$$D_c = \frac{m_c - (n_c - 1)}{n_c(n_c - 1)/2 - (n_c - 1)} \tag{3}$$

This is the number of edges in community c normalized by the minimum and maximum possible number of edges between nodes. Link density D_c is 1 when all member nodes are connected to each other and 0 when edges do not share any nodes (except for the *keystone* node) in a community c. The partition density D for the network, which measures the density or cohesion of edges, is then calculated as the average of link densities D_c as in eq. (4).

$$D = \frac{\sum_c m_c D_c}{\sum_c m_c} \tag{4}$$

4. Labor market areas in the US

To identify LMAs, we apply the link vector community method to county-to-county commuting flow data for 2006-2010 as published by the US Census Bureau.[4] Here, the nodes are counties, while the weighted edges indicate the number of commuters between counties. Our calculations show that the similarity between link vectors is found to follow a power-law[5] distribution (Fig. 5): most pairs of edges have very low similarities while a smaller number of edge pairs have relatively large similarities. The first dot in figure 5 (similarity is 0.11 and probability is 0.05), indicates that 5% of all 262,453 edge pairs have similarity 0.11. Thus, most edge pairs are not related to one another but a few edge pairs have a strong relationship and are classified as belonging to one community. Such pairs with high similarities are candidates for membership in the same LMAs. To formalize the delineation of such areas we next need the concept of a threshold.

[4] Available at http://www.census.gov/hhes/commuting/

[5] This is also means we are dealing with a scale-free network as opposed to a random network.

The similarity at which the maximum partition density *D* occurs is referred to as the 'threshold,' which allows us to sort nodes into communities (lower panel in Fig. 6a). Near the threshold, edges within the communities have maximum partition density *D*; here the network generates the most dense or cohesive structure and the system has well-aggregated subsystems. The threshold for the 2006-2010 US commuting network is 0.5149, which is shown by the red line in the figure. Detected communities with the threshold cover 96.6% of all commuters (upper panel in Fig. 6a).

A total of 6,628 link communities are detected among the 3,143 interrelated US counties, and the size of these communities follows a power law distribution (Fig. 6b). The largest 20 communities account for more than half of all commuters (53.4%) while the smallest 4,137 communities contain only 1% of all commuters. A total of 287 communities have more than ten thousand commuters, and these include 90% of all US commuters and 65.4% of all counties.

In a network, not all nodes connect directly with one another, but they do connect to other communities' nodes indirectly through so-called upper level or core nodes (Ravasz and Barabási, 2003; Jiang and Ishida, 2008). These core or upper level nodes represent not only their own communities but they also form "bridges" between other communities, in a sense creating communities of communities. As the requirement for similarity declines, individual communities grow and absorb new counties; core counties are the most likely to be selected in this growth process (Barabási and Albert, 1999).

Fig. 6c shows how many communities overlap one another in terms of the counties that are contained within each. In the U.S. commuting network, the average county belongs to seven different LMAs or communities (this is the median value in Fig. 6c), but a few counties belong to many communities. In particular, large cities and counties adjacent to large cities

likely belong to many other LMAs because their workers commute to a greater variety of other counties. These LMAs overlap one another through the shared counties. For example, Tarrant County (TX) adjacent to Dallas city (TX) and Gwinnett County (GA) adjacent to Atlanta (GA) belong to more distinct LMAs (or communities) than does any other county.

The link communities method also shows the relationship between a core area (large city) and its surrounding areas (Fig. 7a). The largest LMA in our dataset is the DC-MD-VA area. This area contains 2.5 million commuters (6.9% of all US commuters), 951 edges (2.7% of all edges), and 78 counties. The core county, the District of Columbia (DC), belongs to 55 different overlapping LMAs. The second largest LMA is the NY-NJ area with 2.2 million commuters (6.2%), 368 edges, and 29 counties. The NY-NJ area includes the counties that surround New York County (Manhattan), but it does not contain New York County (NY) itself. Because the residents of New York County (NY) have distinctively different commuting patterns compared with residents of surrounding counties, New York County (NY) forms an independent LMA: it is in fact the third largest. The New York County LMA has only 30 counties and 29 edges (and it thus has a star shape), but 1.7 million commuters (4.7%). The fourth largest LMA is the Boston area with 1.3 million commuters, 131 edges, and 18 counties. The four largest LMAs in the U.S. are all in the Northeast region. Although they are located contiguously and maintain close interchanges with one another, four clearly independent commuting systems exist here, anchored by the District of Columbia (DC), New York (NY) and its surrounding area, and Suffolk (MA), which act as respective centers. Further, these LMAs overlap one another as shown in Fig. 7b, with many counties belonging to more than one LMA. By allowing overlapping LMAs within a hierarchy, we can analyze relationships both within and between interdependent LMAs. The core counties of the next largest LMAs are Fulton (GA), Los Angeles (CA), Cook (IL), San Francisco (CA), and Dallas (TX), respectively.

5. Summary and Conclusion

The most general form of any network has links that are both directed and weighted, as is true of commuting networks. LMAs can be regarded as communities within a commuting network and they can be identified using the community detection method for complex networks. Further, overlapping and hierarchy are two crucial and distinctive features of a community structure. The link community method was introduced by Ahn *et al*. (2010) as a solution that allows for both overlapping and hierarchical community structures, but they stopped short of applying their method to directed and weighted networks. In this study, we adapted the link community method for directed and weighted network using link vectors and applied it to the U.S. commuting network to detect LMAs.

The proposed method identifies community structures in a commuting network for known LMAs, including core-regions and the subtle sub-divisions of the Northeast region. There are two unresolved issues in using link communities as LMAs: isolated counties and small LMAs. Many counties in the Plains and the Rocky Mountain regions are isolated in terms of commuting flows: they have large land areas and most employees work within the counties in which they reside. These counties tend to have independent commuting patterns and to form small LMAs. As noted, the total number of commuters in the smallest 4,137 communities (62.4% of 6,628 total communities) is only 1% of all commuters. We need to develop an additional method for handling isolated counties and small LMAs; for example, we can merge two communities if their number of commuters is below a threshold and the two communities are similar enough to be one community.

In networks, some nodes are more similar and some edges are stronger than others. Structural equivalence (Wasserman and Faust, 1994, p. 347) and Simmelian ties (Krackhardt,

1999) have been used in social network analysis to measure this degree of similarity of nodes and strength of edges, respectively. Conceptually, similarity as defined in eq. (2) measures the structural equivalence between two edges which share a third party node (keystone). We suggest that the idea of link similarity, as proposed by Ahn *et al*. (2010) and extended by us, can be used to integrate the two measures of structural equivalence and Simmelian ties.

In many social networks, edges have properties such as time, distance, gender, age, education, job, hobby, region of residence and work, etc. These edge properties are difficult to show in a single network that contains information only about the direction and weight of connections. Analyzing such communities with multiple edge properties remains challenging (Mucha *et al*., 2010). Relatedly, Expert *et al*. (2011) separated out the effect of space on mobile phone users to elicit additional insights into the underlying complex network. Such an extension would be valuable in the case of commuting networks, but is beyond the scope of the present paper.

If we regard edges as tensors, it may be possible to solve this problem of multiple edge properties. For example, if a network consists of edges with 3 properties, we can express them as a 3rd order tensor. As noted, a network has a matrix form and a matrix consists of vectors. A vector is a 1st tensor while a matrix is a 2nd tensor. By increasing the order of a tensor, we can capture the information of the multi-layered networks in one object (tensor), and use it to detect the community structure of multi-layered networks. We plan to pursue this in future research.

References

1. Ahn, Y.-Y., J.P. Bagrow, and S. Lehmann. 2010. “Link communities reveal multiscale complexity in networks.” *Nature* 466:761-764.

2. Bacher, H. 2000. “A Probabilistic Clustering Model for Variables of Mixed Type.” *Quality & Quantity* 34:223-235.
3. Barabási, A. -L., and R. Albert. 1999. “Emergence of Scaling in Random Networks.” *Science* 285:509-512.
4. Evans, T.S., and R. Lambiotte. 2009. “Line graphs, link partitions, and overlapping communities.” *Physical Review E* 80:016105. doi:10.1103/PhysRevE.80.016105.
5. Evans, T.S., and R. Lambiotte. 2010. “Line graphs of weighted networks for overlapping communities.” *The European Physical Journal B* 77:265-272. doi:10.1140/epjb/e2010-00261-8.
6. Expert, P., T. S. Evans, V. D. Blondel, and R. Lambiotte. 2011. “Uncovering space-independent communities in spatial networks.” *PNAS* 108(19):7663-7668.
7. Florida, R., T. Gulden, and C. Mellander. 2008. “The rise of the mega-region.” *Cambridge Journal of Regions, Economy and Society* 1:459-476. doi:10.1093/cjrse/rsn018.
8. Girvan, M. and, M. Newman. 2002. “Community structure in social and biological networks.” *PNAS* 99:7821-7826.
9. Goetz, S.J. 2014. “Labor market theory and models.” In M.M. Fischer and P. Nijkamp, ed. *Handbook of Regional Science*. Verlag Berlin Heidelberg: Springer, pp.35-57. doi:10.1007/978-3-642-23430-9_6
10. Goetz, S.J., Y. Han, J.L. Findeis, and K.J. Brasier. 2010. “U.S. Commuting Networks and Economic Growth: Measurement and Implications for Spatial Policy.” *Growth and Change* 41(2):276-302.
11. Jiang Y., and T. Ishida. 2008. “Local interaction and non-local coordination in agent social law diffusion.” *Expert Systems with Application* 34:87-95. doi:10.1016/j.eswa.2006.08.019

12. Krackhardt, D. 1999. “The ties that torture: Simmelian tie analysis in organizations.” *Research in the sociology of Organizations* 16:183-210.
13. Lancichinetti, A., S. Fortunato, and J. Kertész. 2009. “Detecting the overlapping and hierarchical community structure in complex networks.” *New Journal of Physics* 11:033015. doi:10.1088/1367-2630/11/3/033015.
14. Milo, R., S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. 2002. “Network Motifs: Simple Building Blocks of Complex Networks.” *Science* 298:824-827. doi:10.1126/science.298.5594. 824.
15. Mucha, P.J., T. Richardson, K. Macon, M.A. Porter, and J.-P. Onnela. 2010. “Community Structure in Time-Dependent, Multiscale, and Multiplex Networks.” *Science* 328:876-878.
16. Newman, M.E.J. 2004. “Fast algorithm for detecting community structure in networks.” *Physical Review E* 69:066133.
17. Palla, G., A.-L. Barabási, and T. Vicsek. 2007. “Quantifying social group evolution.” *Nature* 446:664-667.
18. Palla, G., I. Derenyi, I. Farkas, and T. Vicsek. 2005. “Uncovering the overlapping community structure of complex networks in nature and society.” *Nature* 435:814-818. doi:10.1038/nature/03607.
19. Ravasz, E., and A.-L. Barabási. 2003. “Hierarchical organization in complex networks.” *Physical Review E* 67:026112.
20. Ravasz, E., A.L. Somera, D.A. Mongru, Z.N. Oltvai, and A.-L. Barabási. 2002. “Hierarchical Organization of Modularity in Metabolic Networks.” *Science* 297:1551-1555. doi:10.1126/science.1073374.

21. Shen, H., X. Cheng, K. Cai, and M.-B. Hu. 2009. “Detect overlapping and hierarchical community structure in networks.” *Physica A* 338:1706-1712. doi:10.1016/j.physa.2008.12.021.
22. Tolbert, C.M., and M. Sizer. 1996. *U.S. commuting zone and labor market areas: 1990 update*. Washington DC: U.S. Department of Agriculture, Economic Research Service, Rural Economy Division.
23. Wasserman, S., and K. Faust. 1994. *Social Networks Analysis: Methods and Applications*, New York: Cambridge University Press.
24. Watts, D.J., and S.H. Strogatz. 1998. “Collective dynamics of ‘small-world’ networks.” *Nature* 393:440-442. doi:10.1038/30918.

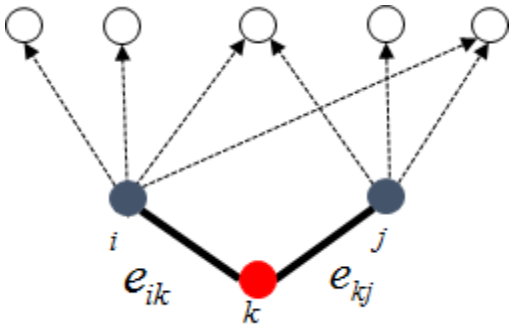

Figure 1 Schematic of a keystone node and link vectors. Keystone k is located between two connected edges e_{ik} and e_{jk}.

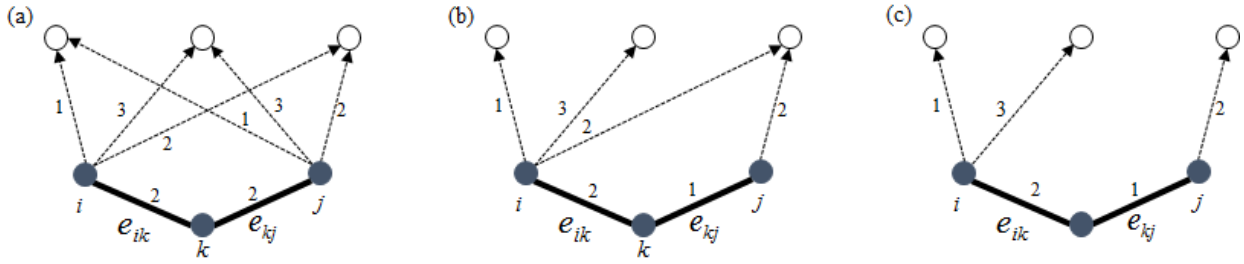

Figure 2 Examples of similarity measure $s_{ij,k}$ between two link vectors $\boldsymbol{e_{ik}}$ and $\boldsymbol{e_{jk}}$. Node k is a keystone and $\boldsymbol{e_{ik}}$ and $\boldsymbol{e_{kj}}$ are link vectors. In the fully-connected (or fully-overlapping) case (a) $\boldsymbol{e_{ik}}\cdot\boldsymbol{e_{jk}}=18$, $|\boldsymbol{e_{ik}}|=|\boldsymbol{e_{jk}}|=4.24$, and $\boldsymbol{s}=1$. In the partially-overlapping case (b) $\boldsymbol{e_{ik}}\cdot\boldsymbol{e_{jk}}=6$, $|\boldsymbol{e_{ik}}|=4.24$, $|\boldsymbol{e_{jk}}|=2.23$, and $\boldsymbol{s}=0.63$. In the isolated case (c) $\boldsymbol{e_{ik}}\cdot\boldsymbol{e_{jk}}=2$, $|\boldsymbol{e_{ik}}|=3.74$, $|\boldsymbol{e_{jk}}|=2.23$, and $\boldsymbol{S}=0.23$.

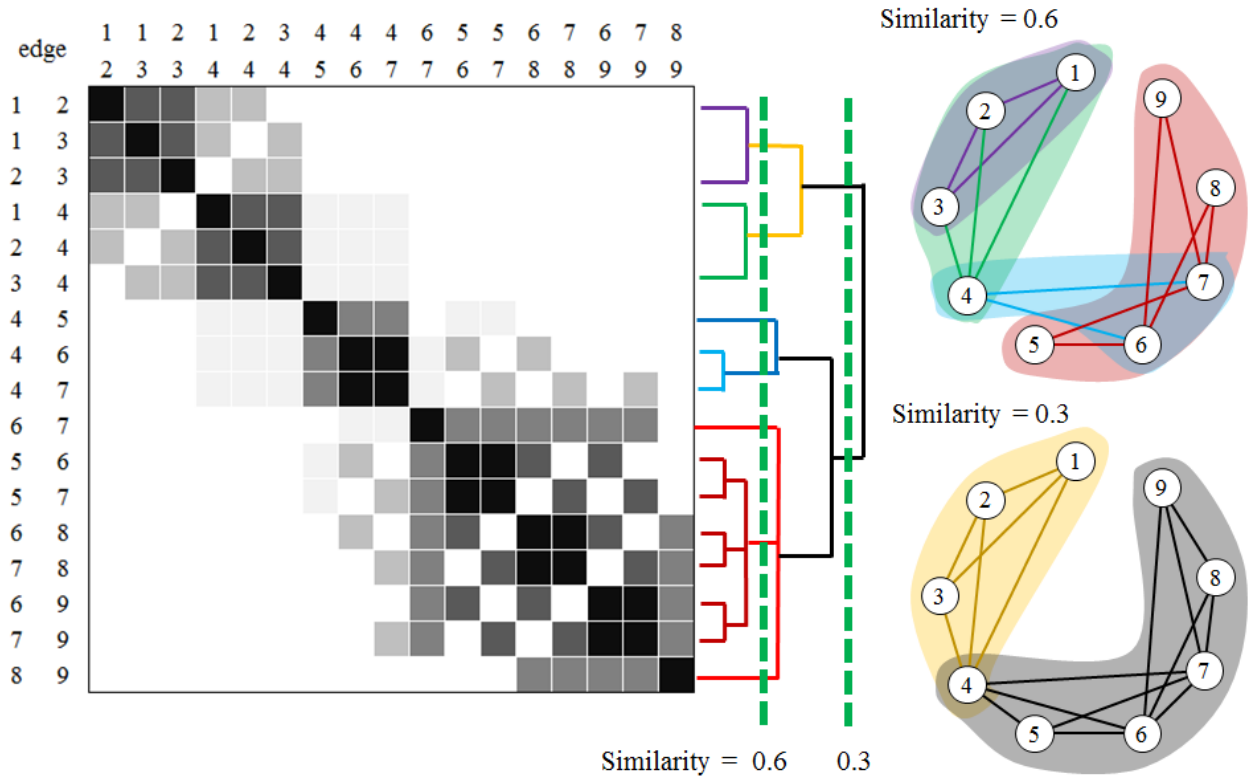

Figure 3 An example of a network and the resulting similarity matrix *s* of link vectors with its link dendrogram. The communities at high (0.6) and low (0.3) threshold are shown. The high threshold yields more communities (four) while the low threshold leaves fewer (two).

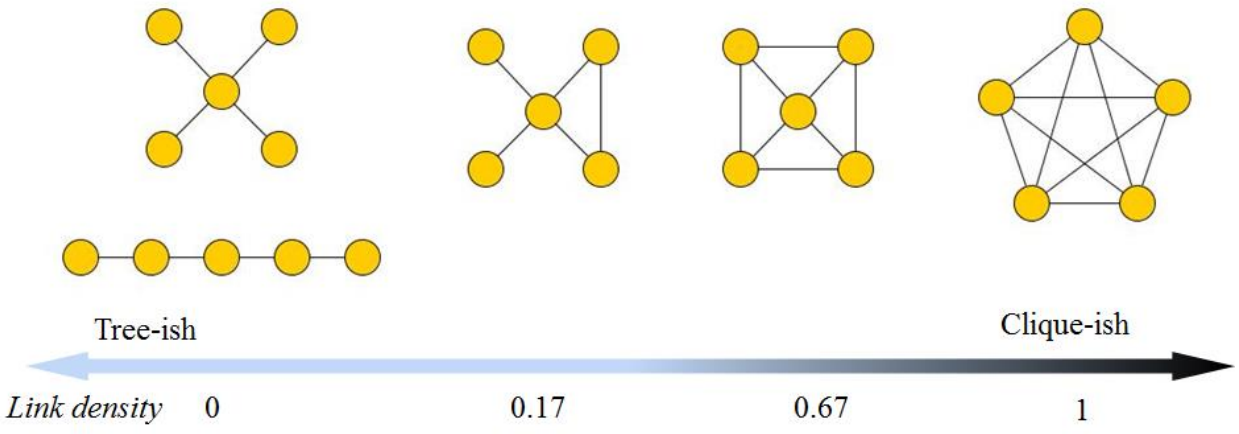

Figure 4 Schematic examples of the *link density* D_c of networks consisting of 5 nodes

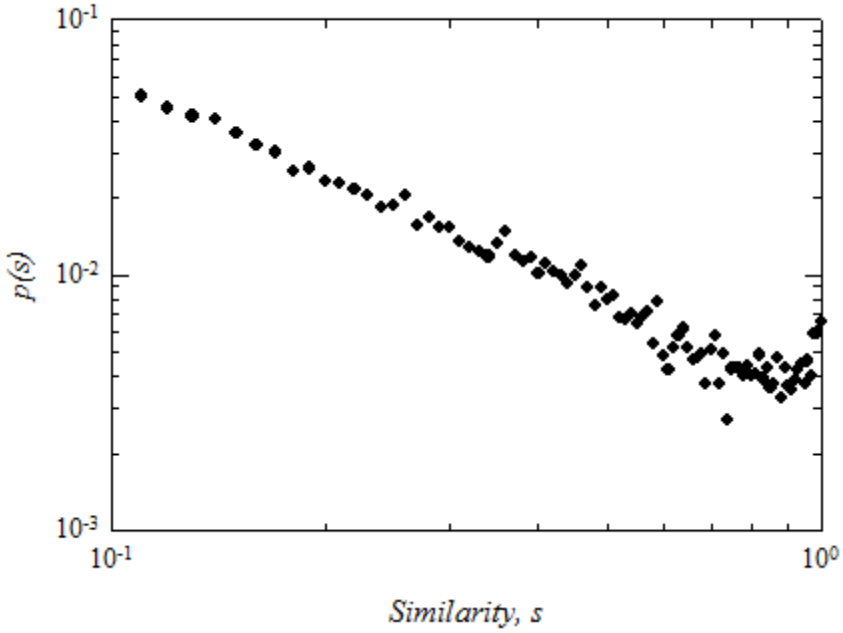

Figure 5 Distribution of similarity *s* between link vectors. The x-axis measures the similarity between links, and the y-axis the probability (or number) of link pairs with the similarity given by the x-axis.

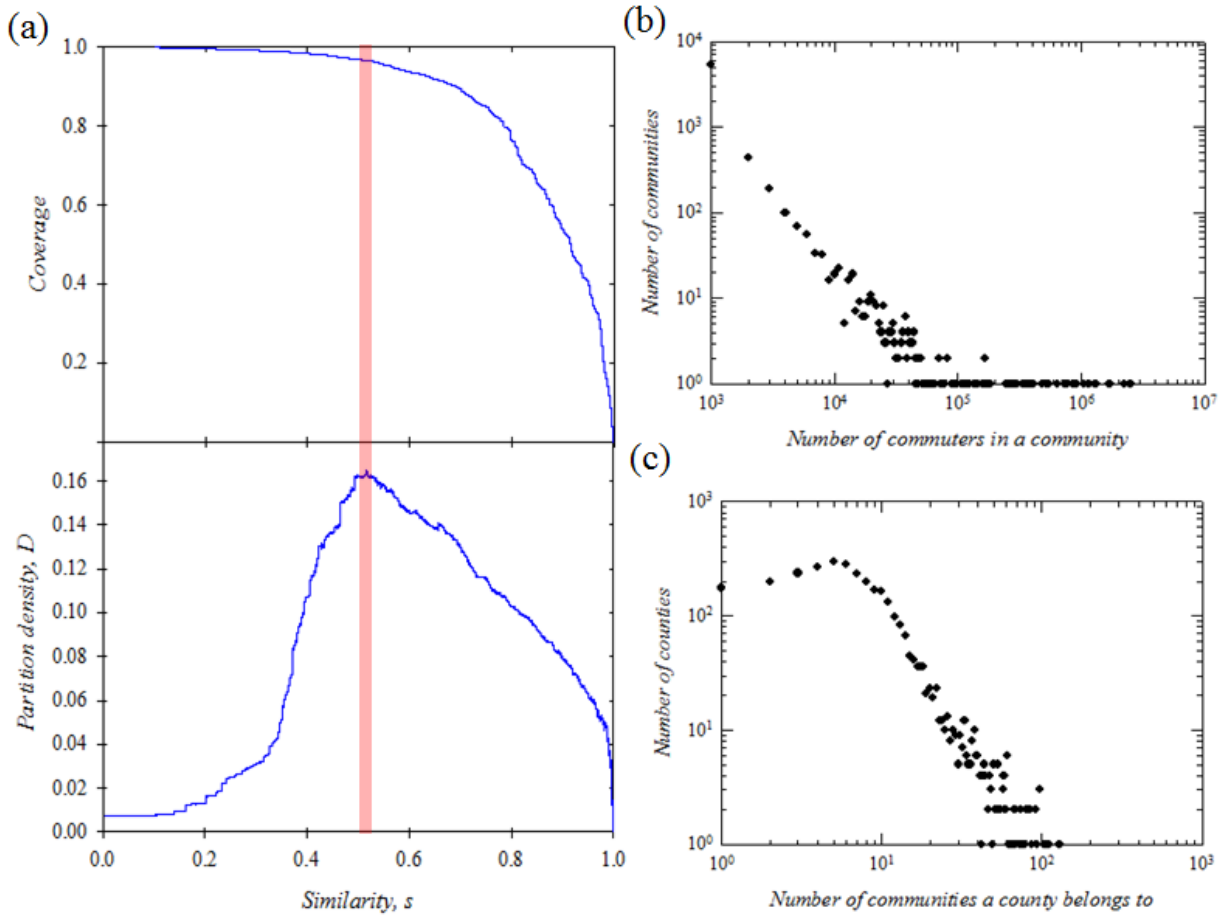

Figure 6. Network statistics for the US commuting network. (a) Summary of the commuting network community structure according to decreasing similarity. Upper part: community coverage (this is the share of all commuters who are classified into commuting communities). Lower part: partition density *D* versus level of similarity *s*. (b) The distribution of community sizes and (c) the number of communities a county belongs to based on a threshold similarity of 0.5159.

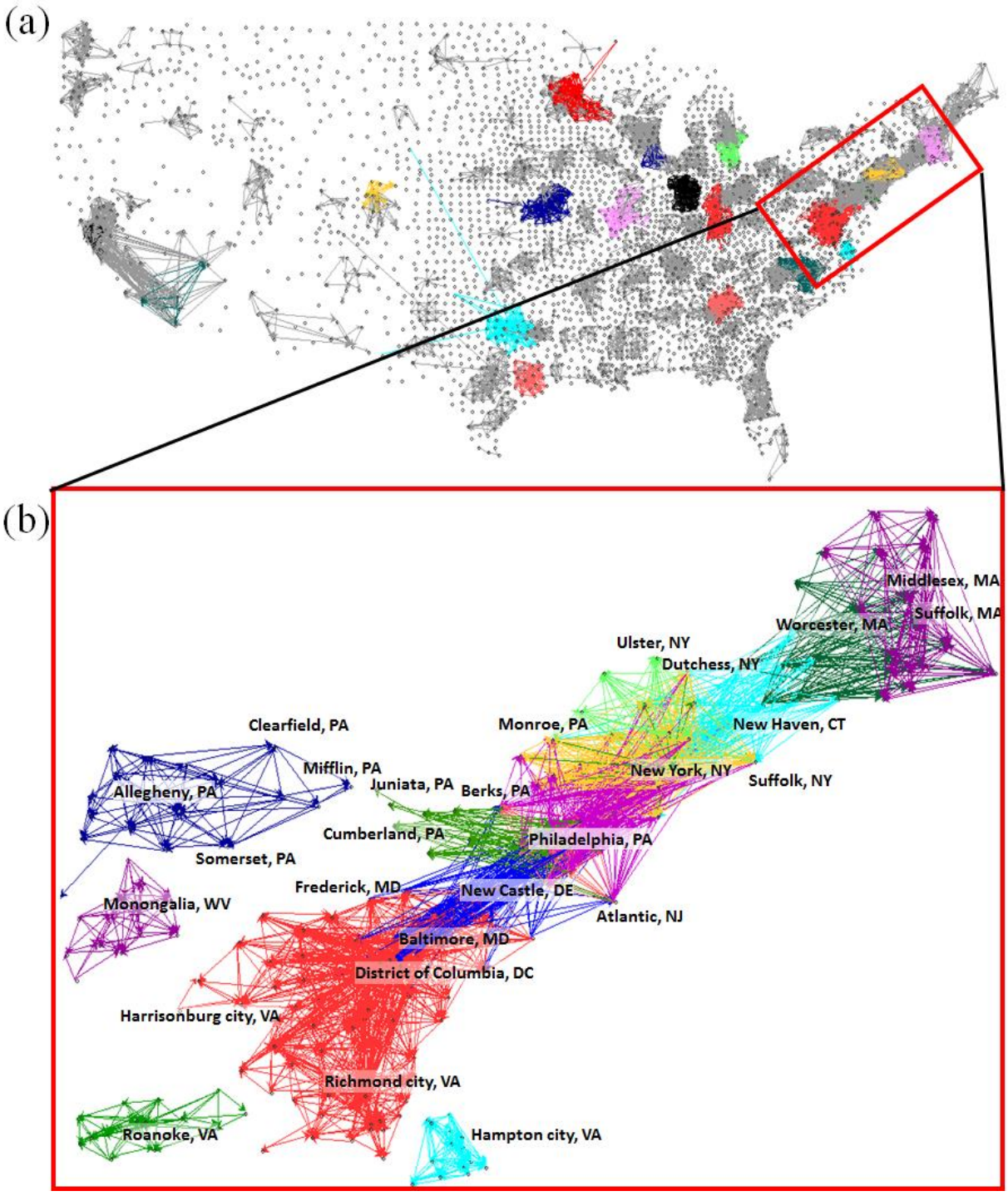

Figure 7 Map of US LMAs as link communities of a commuting network. (a) The map of the largest 287 LMAs (with more than ten thousand commuters). The largest 20 LMAs are colored. (b) The overlapping LMAs of the Northeast region.