Local dominance unveils clusters in networks

Ruiqi Li[1*] **and Xinmiao Sun**[2]

[1]Hangzhou International Innovation Institute, Beihang University, Beijing 100191, China
[2]School of Automation, University of Science and Technology of Beijing, Beijing 100083, China

Abstract

Clusters or communities can provide a coarse-grained description of complex systems at multiple scales, but their detection remains challenging in practice. Community detection methods often define communities as dense subgraphs, or subgraphs with few connections in-between, via concepts such as the cut, conductance, or modularity. Here we consider another perspective built on the notion of local dominance, where low-degree nodes are assigned to the basin of influence of high-degree nodes, and design an efficient algorithm based on local information. Local dominance gives rises to community centers, and uncovers local hierarchies in the network. Community centers have a larger degree than their neighbors and are sufficiently distant from other centers. The strength of our framework is demonstrated on synthesized and empirical networks with ground-truth community labels. The notion of local dominance and the associated asymmetric relations between nodes are not restricted to community detection, and can be utilised in clustering problems, as we illustrate on networks derived from vector data.

More specifically, in this paper, we propose a community detection algorithm in networks, Local Search (LS), that explicitly uses the notion of local dominance and identifies community centres based on local information. In our method, every node is given at most one parent node deemed to be higher up in a partial ranking. Nodes that have a dominant position in their immediate neighborhood[1] or even beyond are identified as local leaders[1]. This defines a rooted tree that spans the network and gives rise to community centers that are *local leaders*[1] with both a larger degree than the nodes in their basin of attraction and a relatively long distance to other local leaders higher up in the ranking. Our approach possesses several interesting properties. Firstly, it provides a new perspective on community detection and delivers community centres and a hierarchy within the community and even a hierarchy among communities as an explicit part of our algorithm, and so mimics advantageous features of the methods based on embedding data in a metric space. Secondly, the identification of communities through local dominance is highly efficient, as it uses purely local topological information and breadth-first search, and runs in linear time. The method does not require the heuristic optimization of an objective function that relies on a global null model[2–7] or computationally costly spreading dynamics[8–10]. Also, our method does not rely on a similarity measure for which there is an wide choice, with an associated uncertainty and variability in results, such as is found in hierarchical clustering based methods[11–14]. Finally, LS is not as susceptible to noise as most methods[15,16], and is less therefore susceptible to finding spurious communities in random graph model realisations[17].

We demonstrate the strength of LS on several classical but challenging synthetic benchmarks and on standard empirical networks with known ground-truth community labels. Our numerical evaluation also includes network representations derived from vector data. As the LS method naturally provides community centres and local hierarchies, it creates an explicit analogy with the notion of cluster centres and distances within clusters that are found in vector clustering methods. Moreover, we also show that applying LS on discretised version of data cloud points outperforms classical unsupervised vector data clustering methods on benchmarks[18].

References

1. Blondel, V. D., Guillaume, J.-L., Hendrickx, J. M., de Kerchove, C. & Lambiotte, R. Local leaders in random networks. *Phys. Rev. E* **77**, 036114 (2008).
2. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008).
3. Clauset, A., Newman, M. E. & Moore, C. Finding community structure in very large networks. *Phys. Rev. E* **70**, 066111 (2004).
4. Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci.* **105**, 1118–1123 (2008).
5. Peel, L., Larremore, D. B. & Clauset, A. The ground truth about metadata and community detection in networks. *Sci. Adv.* **3**, e1602548 (2017).
6. Duch, J. & Arenas, A. Community detection in complex networks using extremal optimization. *Phys. Rev. E* **72**, 027104 (2005).

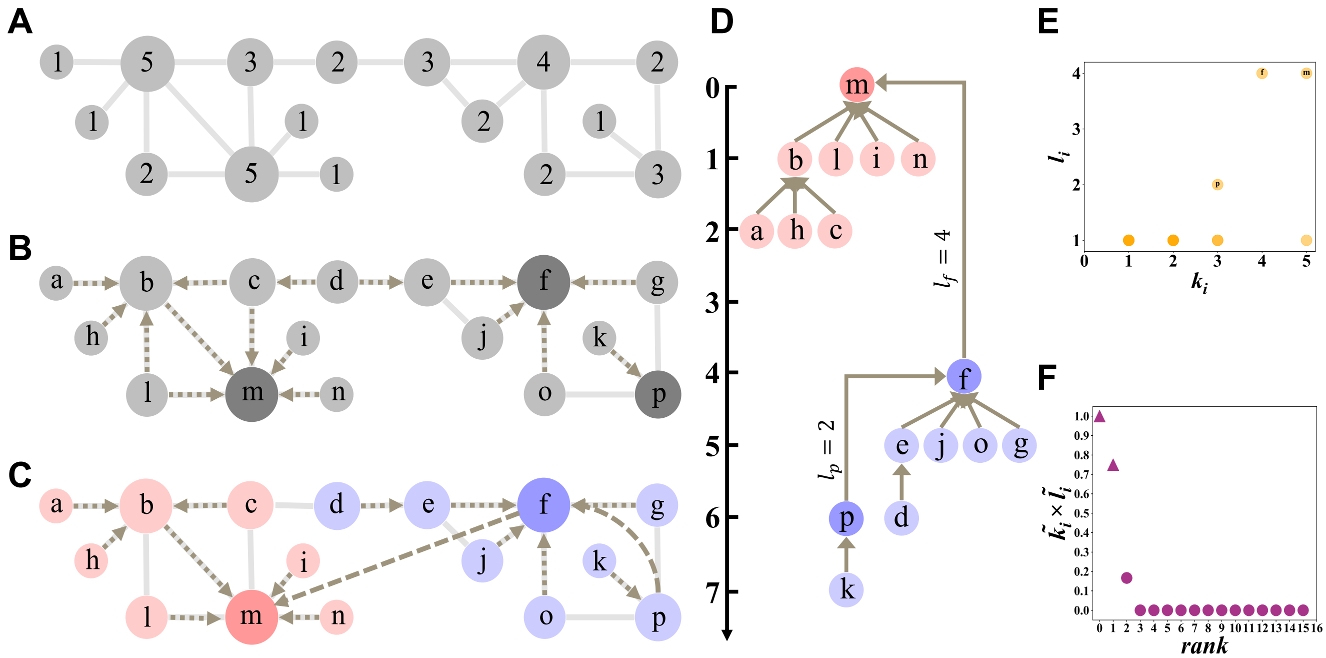

Figure 1. Schematic illustration of the Local Search (LS) algorithm. (**A**) An example network where digits on nodes and size of nodes indicate the degree. (**B**) The identification of local leaders based on local dominance by creating a forest of DAGs as indicated by short dashed directed edges. For each node u, it points to any adjacent neighbor v with $k_v \geq k_u$ and $k_v = \max\{k_z|z \in \mathbf{V}(u)\}$, where $\mathbf{V}(u)$ is the set of neighboring nodes. In this example, nodes are traversed by their lexicographical order, when node b is traversed, it points to m as $k_m = \max\{k_z|z \in \mathbf{V}(b)\} \geq k_b$; later, when m is traversed, it has no out-going link, and so m is identified as a local leader: it does not point to any of its followers and its remaining neighbors all have smaller degrees. When there are more than one neighbor with the same largest degree, more than one directed edge is temporarily added, e.g., node c points to both b and m as $k_b = k_m = \max\{k_z|z \in \mathbf{V}(c)\} \geq k_c$; nodes d and l also have more than one outgoing link. The local leaders, which are potential community centers, are f, m, and p (indicated by dark grey color). (**C**) Each node randomly retains just one out-going edge shown as a short dashed directed edge (e.g., c can point to b or m with an equal probability, similarly for l and d). Then, for each local leader u, a local-BFS is performed to find its nearest local leader with $k_v \geq k_u$, and the shortest path length on network $d_{uv}, \forall v$ is designated by l_u. Here, $p \rightarrow f$ with $l_p = 2$, and $f \rightarrow m$ with $l_f = 4$. In (**C**), short-dash arrows and long-dash arrows correspond to pure followers (whose $l_u = 1$) and local leaders (whose $l_u \geq 2$), respectively. Each node has at most one out-going link $(u \rightarrow v)$, which can go beyond direct connections. The local leader(s) with the maximal degree has no out-going link (here node m). (**D**) The corresponding tree structure formed by local dominance. The scale on the left is a visual aid for calculating l_i between connected nodes in the DAG. (**E**) The scatter plot of k_i and l_i for all nodes. Community centers are of both a larger degree k_i and a longer l_i. (**F**) The decision graph for quantitatively determining community centers (indicated by triangles) based on the product of rescaled degree $\tilde{k}_i$ and rescaled distance $\tilde{l}_i$ (see more details in Supplementary Note 1.2). Community centers can be detected by a visual inspection for obvious gaps or sophisticated automatic detection methods. Here, two centers, nodes m and f, are identified. The color of nodes in (**C**) and (**D**) represents the community partition, and community centers are highlighted by a darker hue of the same color.

7. McLachlan, G. J. & Krishnan, T. *The EM algorithm and extensions*, vol. 382 (John Wiley & Sons, 2007).
8. Raghavan, U. N., Albert, R. & Kumara, S. Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E* **76**, 036106 (2007).
9. Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972–976 (2007).
10. Delvenne, J.-C., Yaliraki, S. N. & Barahona, M. Stability of graph communities across time scales. *Proc. Natl. Acad. Sci.* **107**, 12755–12760 (2010).
11. Girvan, M. & Newman, M. E. Community structure in social and biological networks. *Proc. Natl. Acad. Sci.* **99**, 7821–7826 (2002).
12. Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabási, A.-L. Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555 (2002).
13. Sales-Pardo, M., Guimera, R., Moreira, A. A. & Amaral, L. A. N. Extracting the hierarchical organization of complex systems. *Proc. Natl. Acad. Sci.* **104**, 15224–15229 (2007).
14. Clauset, A., Moore, C. & Newman, M. E. Hierarchical structure and the prediction of missing links in networks. *Nature* **453**, 98–101 (2008).
15. Hoffmann, T., Peel, L., Lambiotte, R. & Jones, N. S. Community detection in networks without observing edges. *Sci. Adv.* **6**, eaav1478 (2020).
16. Fortunato, S. Community detection in graphs. *Phys. Reports* **486**, 75–174 (2010).
17. Reichardt, J. & Bornholdt, S. When are networks truly modular? *Physica D* **224**, 20–26 (2006).
18. Rodriguez, A. & Laio, A. Clustering by fast search and find of density peaks. *Science* **344**, 1492–1496 (2014).