Map Equation Centrality: A Map Equation-based Community-Aware Centrality Score

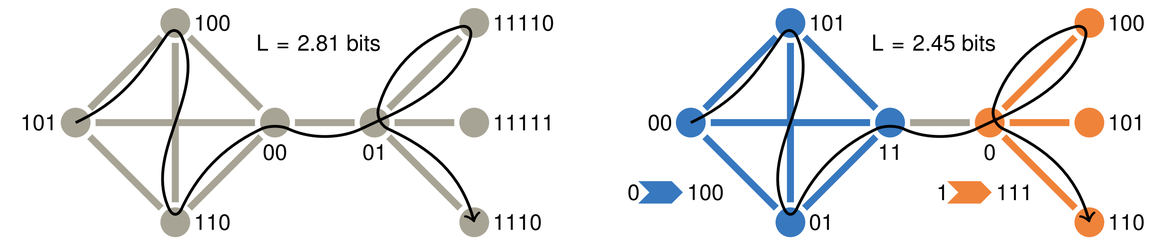
Christopher Blöcker[1], Juan Carlos Nieves Sanchez[2], and Martin Rosvall[1]

[1]Integrated Science Lab, Department of Physics, Umeå University
[2]Department of Computing Science, Umeå University

Abstract

To measure node importance, network scientists employ centrality scores that typically take a microscopic or macroscopic perspective, relying on node features or global network structure. However, traditional centrality measures, such as degree centrality and PageRank, neglect the community structure found in real-world networks. To study node importance based on network flows from a mesoscopic perspective, we exploit the coding principles behind the map equation framework, and derive a community-aware information-theoretic centrality score analytically. Applied to artificial and real-world networks, we demonstrate that our approach enables a more fine-grained differentiation between nodes than node-local or network-global measures, and highlight the role that local network context plays in determining node importance.

The Map Equation Framework



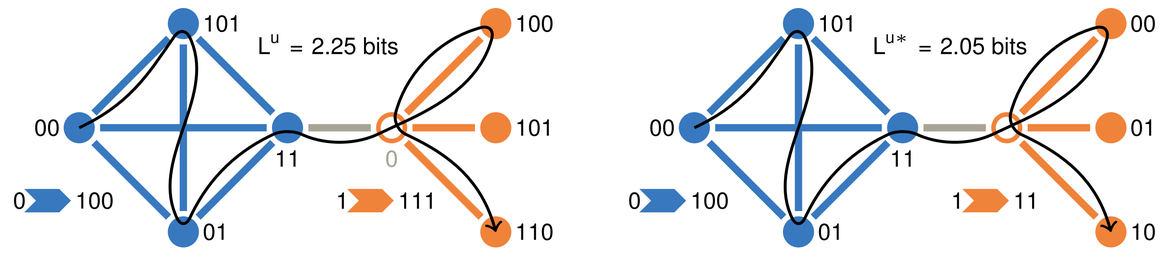
(a) A one-level partition with unique codewords for each node.

(b) A modular partition with unique codewords for nodes within modules.

Figure 1: Coding example for a small network with eight nodes and two communities. (a) In the one-level partition where each node has a unique codeword, the codelength is 2.81 bits. (b) In a modular partition, the codelength is reduced because codewords can be re-used between modules, reducing the codelength to 2.45 bits.

- The map equation is an information-theoretic objective function for community detection
- It measures the quality of a network partition by relating it to the lower bound of the average per-step description length for a random walk
- A network partition corresponds to a modular coding scheme, based on a Huffman code

Map Equation Centrality



(a) Using the same coding scheme as before, but without using the codeword for the node shown as a circle.

(b) Designing a new coding scheme where the node shown as a circle has no codeword.

Figure 2: Effect of silencing a node, that is, not encoding for random-walker steps to the respective node. (a) Using the same coding scheme, but simply omitting the node shown as a circle from the random walk's description. This is inefficient because the codeword assigned to the node is never used. (b) Designing a new coding scheme that does not assign a codeword to the node shown as a circle. The rest of the nodes receive shorter codewords.

- To measure node importance, we consider by how much more the random walk's description can be compressed if the node was not present
- By not assigning a codeword to a node, on average, the rest of the nodes can receive shorter codewords
- Essentially, a node's importance is the effective marginal harm it causes to other nodes by its existence

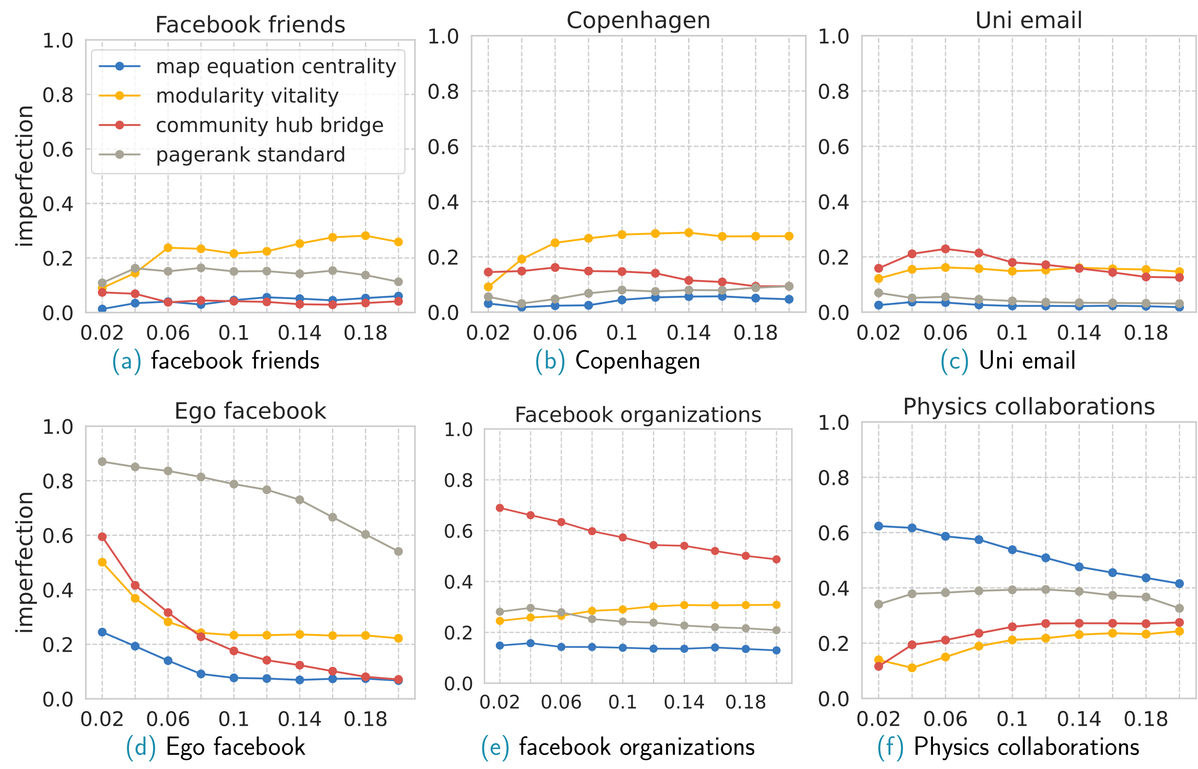
Results

- We have tested Map Equation Centrality on a set of online social networks and compared it with other community-aware centrality measures
- To measure the performance of centrality scores, we test how well they identify nodes with high spreading power
- Spreading power is the expected number of nodes that get infected in an SIR disease spreading model where the disease starts at the node in question, and nodes recover after 1 time step

Results

Table 1: List of networks and their properties: number of nodes (N), number of links ($|E|$), number of communities as detected with infomap (M), mixing (μ), and epidemic threshold (λ).

Network	N	$\|E\|$	M	μ	λ
facebook friends	329	1,954	20	0.127	0.048
Copenhagen	800	6,429	36	0.502	0.038
Uni email	1,133	5,452	52	0.402	0.057
Ego facebook	4,039	88,234	74	0.082	0.009
facebook org.	5,524	94,219	48	0.352	0.016
Physics collab.	8,798	27,416	863	0.279	0.066



(a) facebook friends (b) Copenhagen (c) Uni email

(d) Ego facebook (e) facebook organizations (f) Physics collaborations

Figure 3: Results for three community-aware centrality scores as well as standard PageRank on six online social networks.

- In general, Map Equation Centrality outperforms the other community-aware centrality measures
- However, it is no silver bullet: on the physics collaborations network, the other measures, including PageRank perform better

Conclusion

- We derived a community-aware centrality score from the map equation to measure node importance in modular networks
- In most cases we tested, Map Equation Centrality outperforms other centrality measures

Acknowledgements

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

ArXiv

arXiv:2201.12590

christopher.blocker@umu.se