Path-sharing: A new betweenness measure for community identification in networks

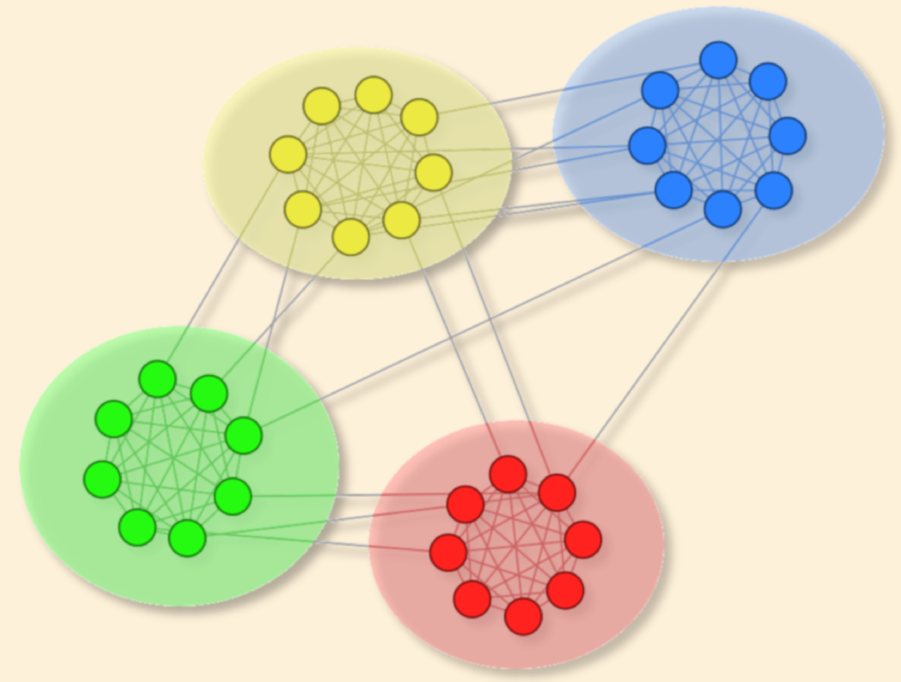
Paul McCarthy
University of Otago, Dunedin, New Zealand

Introduction

Community identification in networks has a wide range of practical applications, including data clustering and social network analysis. We present *path-sharing*, a new measure of betweenness, for use in identifying densely connected clusters in networks. We show that path-sharing performs well at identifying communities in artificial benchmark networks, giving performance comparable to that of state-of-the-art community identification techniques. We also demonstrate a practical use of path-sharing when used in community identification, by applying it to an image segmentation problem.

The problem

Our aim is to identify communities, or densely connected groups of nodes, in a graph. How can we decide which edges lie within communities, and which edges lie between?



We need a measure which tells us the importance of each edge in connecting groups of nodes. Answer: path-sharing.

Path-sharing

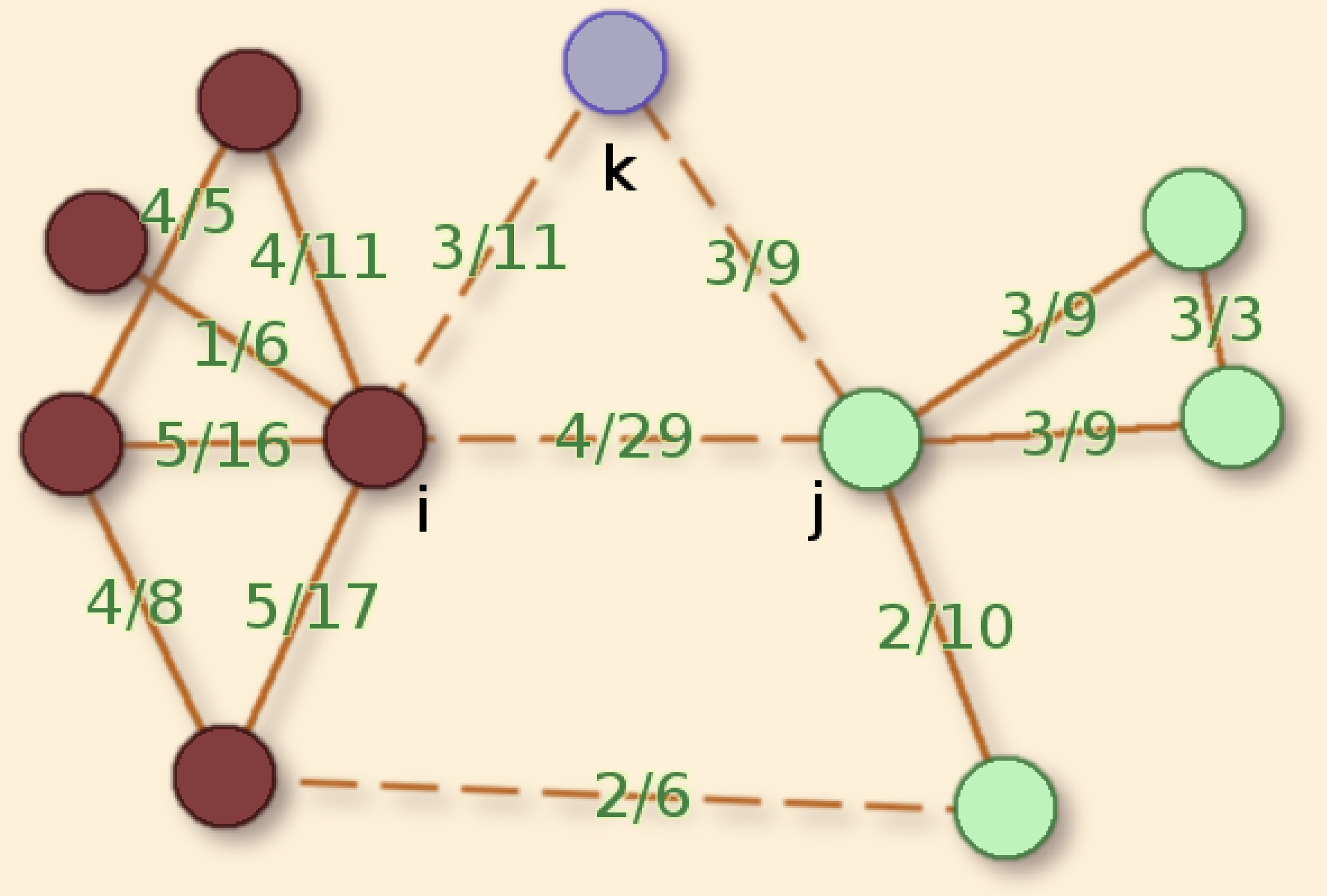
For an undirected graph $G = (V, E)$, the path-sharing P_{ij} for an edge $(i, j, w) \in E$, between vertices $i, j \in V$, with weight w, is calculated as follows:

$$V_i = \{i\} \cup \{u | u \in V, u \neq j, (i, u, w) \in E\}$$
$$V_j = \{j\} \cup \{u | u \in V, u \neq i, (j, u, w) \in E\}$$
$$V_{i \cap j} = V_i \cap V_j$$
$$E_{V_i V_j} = \{(u, v, w) | u \in V_i, v \in V_j, (u, v, w) \in E\}$$
$$W_{V_i V_j} = \sum w | (u, v, w) \in E_{V_i V_j}$$

$$P_{ij} = \frac{W_{V_i V_j}}{|V_i| . |V_j| - |V_{i \cap j}|}$$

Example

Consider this unweighted graph (i.e. all edge weights are equal to 1.0):



The red (darker) vertices (vertex i and its neighbours) form the set V_i, whereas the green (lighter) vertices (vertex j and its neighbours) form the set V_j. Vertex k is a member of both V_i and V_j, hence $V_{i \cap j} = \{k\}$. The path-sharing P_{ij} for the edge which lies between vertices i and j is calculated simply by summing the weights of all edges which lie between the sets V_i and V_j, (the dashed lines). This sum is then divided by the total weight of all such possible edges. In this example, the summed weight of all edges which lie between V_i and V_j is 4, and the total possible weight is $(6 \times 5) - 1$ (1 is subtracted to account for vertex k), giving $P_{ij} = \frac{4}{29} \approx 0.138$.

Community identification

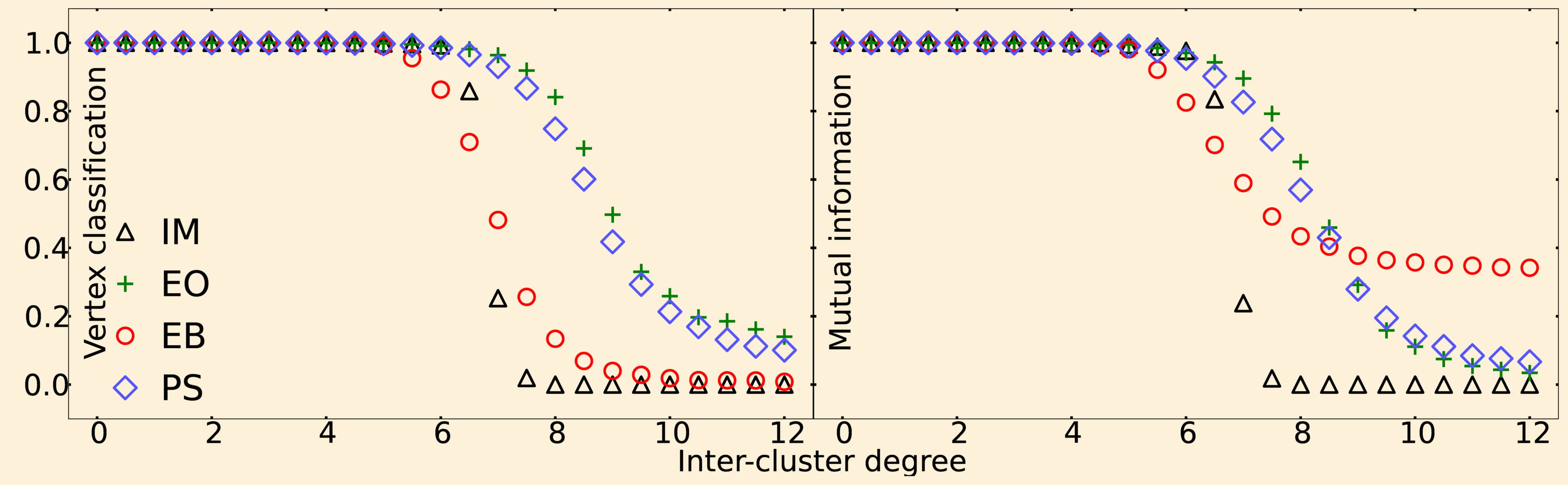
We have adapted Girvan and Newman's divisive clustering algorithm. Start with the graph under analysis, with all edges intact, and proceed as follows:

1. Calculate path-sharing for every edge in the graph.
2. Remove the edge with the lowest path-sharing value. If multiple edges have the same value, randomly select one for removal.
3. Calculate Newman's modularity measure on the original graph, with communities defined by the components created in the current graph. Stop when the maximum possible modularity value has been reached. Otherwise, continue from step 1.

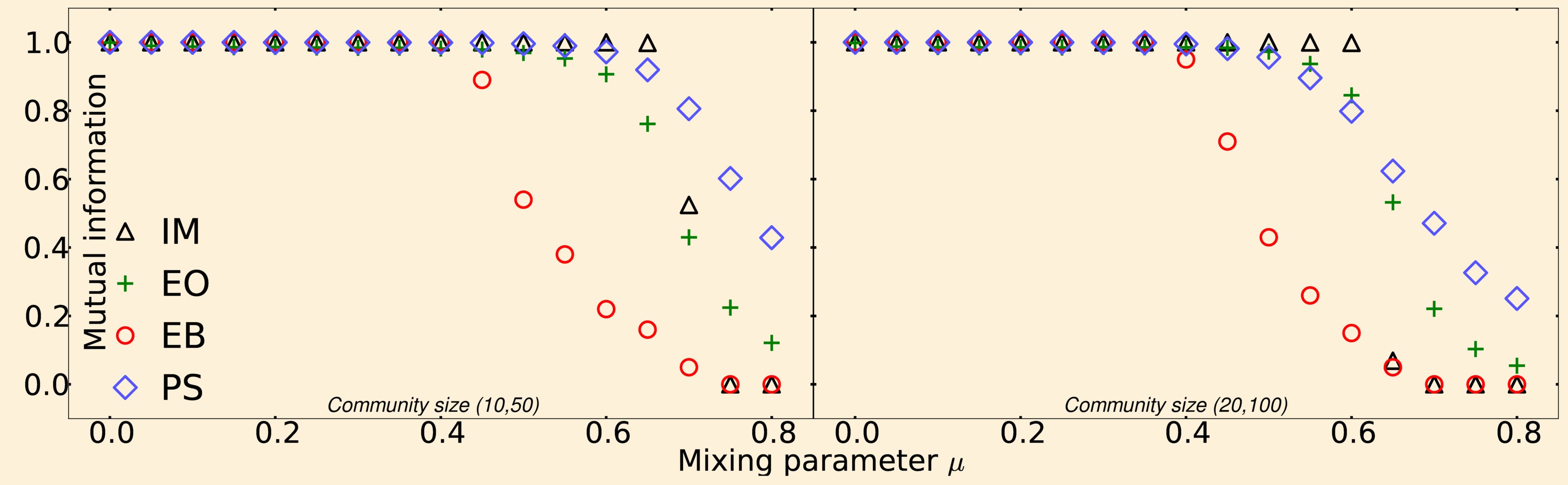
Benchmarks

We compared path-sharing (PS) with three other community identification techniques: Girvan and Newman's edge-betweenness (EB); Duch and Arenas' extremal optimization (EO); and Rosvall and Bergstrom's Infomap (IM).

Girvan and Newman's benchmark networks



Lancichinetti et. al.'s benchmark networks

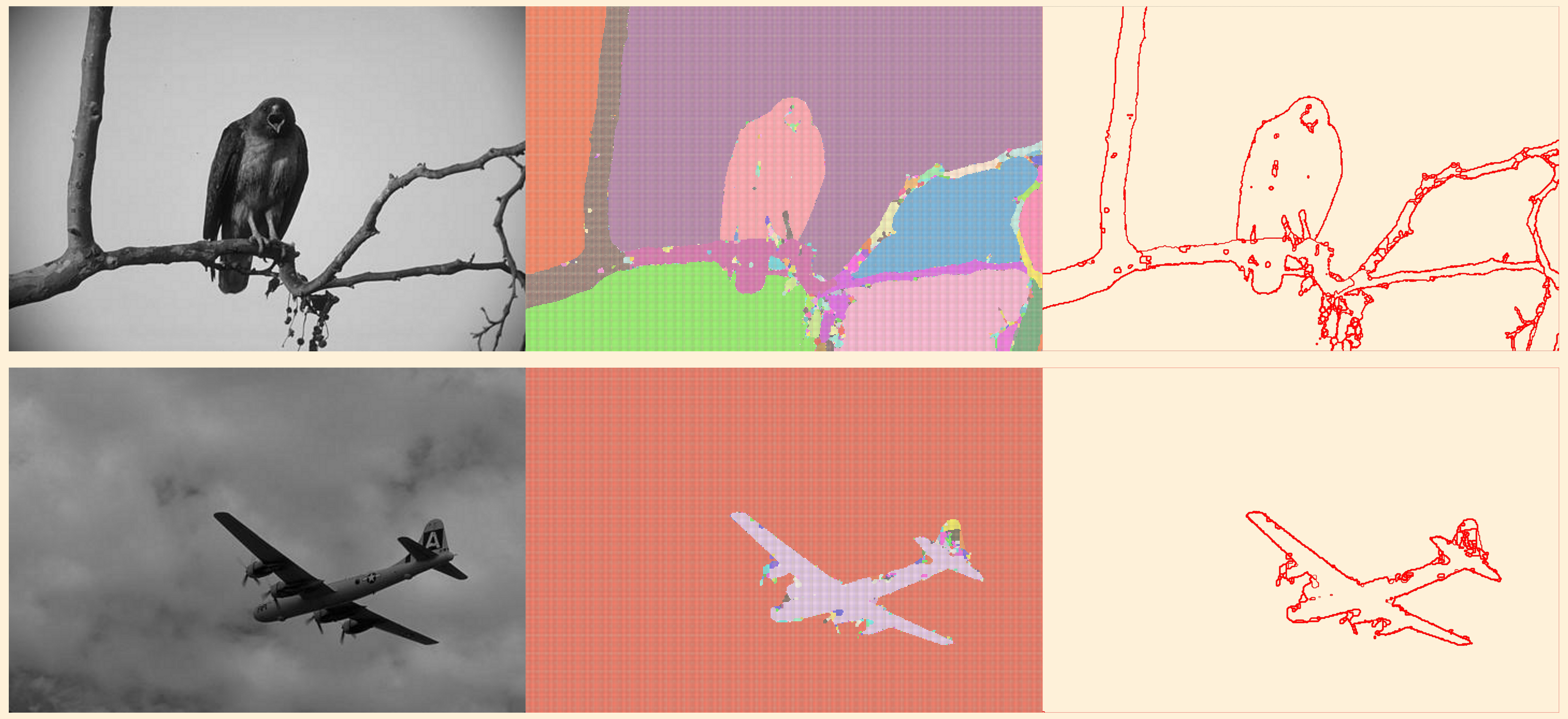


Application: image segmentation

Create an undirected graph from an image using Shi and Malik's approach: every pixel in the image is added as a vertex in a graph $G = (V, E)$; edge weights w_{ij} between each pair of pixels $i, j \in V$ are defined by:

$$w_{ij} = e^{\frac{-|F(i) - F(j)|^2}{\sigma_F^2}} \times \begin{cases} e^{\frac{-|X(i) - X(j)|^2}{\sigma_X^2}} & \text{if } |X(i) - X(j)| < r \\ 0 & \text{otherwise} \end{cases}$$

where: $|F(i) - F(j)|$ is the difference in intensity between pixels i and j; $|X(i) - X(j)|$ is the distance between pixels i and j; σ_F is the intensity scaling parameter (25.0); σ_X is the distance scaling parameter (6.0); and r is the cutoff distance (1.5).



http://miracle.otago.ac.nz/postgrads/pmccarthy/
pauldmccarthy@gmail.com