Community detection and proximity measures
Maximilien Danisch1,2, Jean-Loup Guillaume3 and Bénédicte Le Grand2
1 LIP6 (CNRS – UPMC) and 2 CRI (Université Paris 1)

Context: in complex networks, there are communities

<table>
<thead>
<tr>
<th>networks</th>
<th>nodes</th>
<th>edges</th>
<th>communities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wikipedia</td>
<td>Wikipedia pages profiles</td>
<td>hypertext links file exchanges</td>
<td>Wikipedia categories colleagues, families, sports club communities of interests</td>
</tr>
</tbody>
</table>

**Problematics:** How to find these communities?

- The greedy optimization of a quality function suffers from:
  1. local minima
  2. hidden scale parameters.
- Let’s look for an alternative approach!

**Idea 1:** use the related notion of proximity measure

Rank nodes according to their proximity to a given node.

Irregularities in the decrease can indicate the presence of communities:

- Often a powerlaw is obtained = no scale can be extracted = the node belongs to several communities of various sizes.

**Idea 2:** use the notion of multi-ego-centered community

- While one node generally belongs to many communities, a small set of nodes, e.g., 2, is enough to define a single community.

**Idea 3:** a proximity measure has to be parametrized

- What to do with HUBs?

\[
\text{Pro}_{\alpha,\beta}(i,j) = \frac{1}{d_j^\alpha} N(i,j)
\]

\(N(i,j)\): number of paths of length \(l\) between \(i\) and \(j\), \(d_j\): degree of node \(j\).

- If we know some nodes that are near to one another, e.g., nodes belonging to a same community, we can learn the parameters that rank these nodes as close as possible to one another.

**Application 1:** find all communities of a node

A- Choose a set of candidates:

B- Find bi-ego-centered communities:

C- Clean and label the communities:

- We can use a similar framework to find all overlapping communities in a network.

**Application 2:** community completion

Average number of paths of length 1 to 6 between two nodes in the same community and two nodes in different communities:

Learning a proximity measure, combining scorings and unfolding the community: