Graph Partitioning – Community Detection

① Hierachial Comm. Det.

↳ Modularity

② Spectral Comm. Det.

HCD: Agglomerative / divisive approaches

Q: Choice of similarity? Lots of choices

$O(N_v^2 \log N_v)$ time

Modularity based methods

$$\text{mod}(C) = \sum_{k=1}^{K} \left[f_{kk}(C) - f_{kk}^{*} \right]^2$$

fraction of edges in G connecting vertices in C_i with C_j

expectation under a random graph

Fortunato (Resolution Limits 2007)

– Mod maximization can fail to resolve clusters smaller than certain size depending on N_E

$$\Delta \text{mod}(\cdot)_{ij} = \frac{e_{ij}}{N_e} - 2\left(\frac{d_i}{2N_e}\right)\left(\frac{d_j}{2N_e}\right)$$

$$= \begin{cases} > 0 & : e_{ij} > \frac{d_i d_j}{2N_e} \\ \leq 0 & \text{o.w.} \end{cases}$$

d_i = total degree in C_i $e_{ij} = |E(C_i, C_j)|$

Rmk: doesn't rely on internal structure of clusters/modules.

Good, Montjoye, Clauset (2010)

Spectral Partitioning

- Two types of matrices we consider A, L

- Suppose G is composed of roughly two d-regular graphs

Facts:

1. Largest two eigenvalues will be roughly equal to d.

2. Remaining eigenvalues $O(\sqrt{d})$

Rmk: Generates a spectral gap.

- Signs of entries of top two eigenvectors speak to which community each node belongs to.

<u>Rmk</u>: Eigenvalues of A tend to mirror degree distributions

Spectral Partitioning: Laplacian Matrix

$L = D - A \qquad x^T L x = \sum_{(i,j)\in E} (x_i - x_j)^2$

Thrm: A graph G will consist of K communities iff $\lambda_K(L) > 0 = \lambda_{K+1}(L)$

Isoperimetric number

$$\phi(S,\bar{S}) = \frac{|E(S,\bar{S})|}{|S|} \quad \text{ratio cut} \quad \phi(G) = \min_{S\subset V : |S| < \frac{N_V}{2}} \phi(S,\bar{S})$$

$$\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{\lambda_2(2d_{max} - \lambda_2)}$$

$\phi(G)$ small iff λ_2 small.

Fiedler's Method

$S = \{v \in V : x_2 \geq 0\} \quad \bar{S} = \{v \in V : x_2(v) < 0\}$

$$\phi(G) \le \phi(S,\bar{S}) \le \frac{\phi^2(G)}{d_{max}} \le \lambda_2$$

Rmk: spectral bisection computationally efficient approximation to achieving best cut $\phi(G)$.

Only need top few eigenvectors $\Longrightarrow$ Lanczos can be used $O\left(\frac{N_E}{\lambda_3-\lambda_2}\right)$

- linear in number of edges