L24: Graphs → Communities

April 15, 2020

Jeff M. Phillips
Posters

Turn in PDF in Canvas

2 feet

Title

Introduction

Problem

Set up

Chord

Figure

Thanks
Communities in Graphs

Graph \( G = (V, E) \)

- people/entities
- how edges are defined
- modeling choice
Early web business (circa 90s)

Sociology "social network"

manually collect

100 - 700

Seniors in high school

friends
Mathematical Models of large graphs.

Erdős–Rényi: Pick $n = |U|$, choose $p \in (0,1)$ \implies every edge exists \begin{align*} \quad \vdots & \quad \vdots \end{align*}
Who is more likely to join community?

all friends are connected

more likely

community

4 no friends are connected
Preferential Attachment

- More likely to form edge (a,c) than (a,d)

Average person on Twitter followed by \( \approx 500 \)

Most popular (Obama)

130 million followers
How do we find communities (spectral clustering)?

Important nodes/edges

PageRank

Community = tightly connected subgraph \( V' \subseteq V \)

all edges defined by \( V' \)
Betweenness: important edges.

\[
\text{betweenness} = \frac{\text{fraction of all shortest paths that use edge } (a, b)}{}
\]

- remove important edges
- remaining connected components
- communities

\[
\text{betweenness}(a, b) = \frac{\text{fraction of all shortest paths that use edge } (a, b)}{}
\]
Modularity $G = (V, E)$

Score $Q(C) < cV$

$\equiv \left( \text{fraction of edges in } G_C \right)$

$= \frac{1}{|E_1|} \sum_{i,j \in E_1} A_{ij}$

(expected fraction of edges)

$= \frac{1}{|E_1|} \sum_{i \in C} \sum_{j \in C} E_{ij}$

(adjacency matrix $A = A_{ij} = 1$ if $e_{ij} \in E$)

$\equiv \frac{d_i d_j}{2 |E_1|}$

$= d_i = \text{degree of } v_i$
\[ Q(c) = \frac{1}{4|E|} \sum_{e \in E} \left( A_{ij} - \frac{\delta_e(i,j)}{2} \right) \]

Find communities \( C \subset V \) with largest \( Q(c) \).
Cliques is set of vertices \( C \subseteq V \) where all edges \( ij,j \in C \) have edge \((i,j) \in E\).

A priori Alg.
- 2-cliques start w/ all edges
- 3-cliques Find all triples of edges which have same 3 vertices
- 4-cliques - use 3-cliques, expand greedily