1-means clustering and conductance

Twan van Laarhoven

Radboud University Nijmegen
Open Universiteit Nederland

November 27th, 2017

Radboud University

Outline

Outline

Network community detection with conductance

The relation to k-means clustering

Algorithms

Experiments

Conclusions

Network community detection

Global community detection

Given a network, find all tightly connected sets of nodes (communities).

Local community detection

Given a network and a seed node, find the community/communities containing that seed. Without inspecting the whole graph.

Network community detection

Global community detection

Given a network, find all tightly connected sets of nodes (communities).

Local community detection

Given a network and a seed node, find the community/communities containing that seed. Without inspecting the whole graph.

Communities as optima

Graphs

$G = (V, E)$,
$a_{ij} = a_{ji} = 1$ if $(i, j) \in E$ else 0

Score function

$\phi_G : \mathcal{C}(G) \to \mathbb{R}$

Note: I'll consider sets and vectors interchangeably, so $\mathcal{C}(G) = \mathcal{P}(V)$ or $\mathcal{C}(G) = \mathbb{R}^V$.

Conductance

Radboud University

Definition

Fraction of incident edges leaving the community

$$\phi(c) = \frac{\#\{(i,j) \in E \mid i \in c, j \notin c\}}{\#\{(i,j) \in E \mid i \in c, j \in V\}},$$

or

$$\phi(c) = 1 - \frac{\sum_{i,j \in V} c_i a_{ij} c_j}{\sum_{i,j \in V} c_i a_{ij}}$$

where $c_i \in \{0, 1\}$.

Very popular objective for finding network communities.

Conductance

Definition

Fraction of incident edges leaving the community

$$\phi(c) = \frac{\#\{(i,j) \in E \mid i \in c, j \notin c\}}{\#\{(i,j) \in E \mid i \in c, j \in V\}},$$

or

$$\phi(c) = 1 - \frac{\sum_{i,j \in V} c_i a_{ij} c_j}{\sum_{i,j \in V} c_i a_{ij}}$$

where $c_i \in \{0, 1\}$.

Very popular objective for finding network communities.

Conductance

Definition

Fraction of incident edges leaving the community

$$\phi(c) = \frac{\#\{(i,j) \in E \mid i \in c, j \notin c\}}{\#\{(i,j) \in E \mid i \in c, j \in V\}},$$

or

$$\phi(c) = 1 - \frac{\sum_{i,j \in V} c_i a_{ij} c_j}{\sum_{i,j \in V} c_i a_{ij}}$$

where $c_i \in \{0, 1\}$.

Very popular objective for finding network communities.

Continuous optimization

Radbound University

As an optimization problem

$$\begin{aligned} \underset{c}{\text{minimize}} \quad & \phi(c) \\ \text{subject to} \quad & c_i \in \{0,1\} \quad \text{for all } i. \end{aligned}$$

Karush-Kuhn-Tucker conditions

c is a local optimum if for all c_i

$$0 \leq c_i \leq 1, \text{ and}$$

$$\frac{\partial \phi(c)}{\partial c_i} \leq 0 \quad \text{if } c_i \leq 1, \text{ and}$$

$$\frac{\partial \phi(c)}{\partial c_i} \geq 0 \quad \text{if } c_i \geq 0.$$

As an optimization problem

$$\begin{aligned} \underset{c}{\text{minimize}} \quad & \phi(c) \\ \text{subject to} \quad & 0 \le c_i \le 1 \quad \text{for all } i. \end{aligned}$$

Karush-Kuhn-Tucker conditions

c is a local optimum if for all c_i

$$0 \le c_i \le 1, \text{ and}$$

$$\frac{\partial \phi(c)}{\partial c_i} \le 0 \quad \text{if } c_i \le 1, \text{ and}$$

$$\frac{\partial \phi(c)}{\partial c_i} \ge 0 \quad \text{if } c_i \ge 0.$$

Continuous optimization

As an optimization problem

$$\begin{aligned} \underset{c}{\text{minimize}} \quad & \phi(c) \\ \text{subject to} \quad & 0 \le c_i \le 1 \quad \text{for all } i. \end{aligned}$$

Karush-Kuhn-Tucker conditions

c is a local optimum if for all c_i

$$\begin{aligned} & 0 \le c_i \le 1, \text{ and} \\ & \frac{\partial \phi(c)}{\partial c_i} \le 0 \quad \text{if } c_i \le 1, \text{ and} \\ & \frac{\partial \phi(c)}{\partial c_i} \ge 0 \quad \text{if } c_i \ge 0. \end{aligned}$$

Continuous optimization

As an optimization problem

$$\begin{aligned} \underset{c}{\text{minimize}} \quad & \phi(c) \\ \text{subject to} \quad & 0 \leq c_i \leq 1 \quad \text{for all } i. \end{aligned}$$

Karush-Kuhn-Tucker conditions

c is a local optimum if for all c_i

$$\begin{aligned} &0 \leq c_i \leq 1 \\ &\nabla\phi(c)_i \geq 0 \quad \text{if } c_i = 0 \\ &\nabla\phi(c)_i = 0 \quad \text{if } 0 < c_i < 1, \\ &\nabla\phi(c)_i \leq 0 \quad \text{if } c_i = 1. \end{aligned}$$

Local optima

Radboud University

Local optima are discrete

If c as a strict local minimum of ϕ, then $c_i \in \{0,1\}$ for all i.

Proof sketch

Look at ϕ as a function of a single c_i:

$$\phi(c_i) = \frac{\alpha_1 + \alpha_2 c_i + \alpha_3 c_i^2}{\alpha_4 + \alpha_5 c_i}.$$

If $0 < c_i < 1$ and $\phi'(c_i) = 0$, then $\phi''(c_i) = 2\alpha_3/(\alpha_4 + \alpha_5 c_i)^3 \geq 0$.

Local optima

Local optima are discrete

If c as a strict local minimum of ϕ, then $c_i \in \{0, 1\}$ for all i.

Proof sketch

Look at ϕ as a function of a single c_i:

$$\phi(c_i) = \frac{\alpha_1 + \alpha_2 c_i + \alpha_3 c_i^2}{\alpha_4 + \alpha_5 c_i}.$$

If $0 < c_i < 1$ and $\phi'(c_i) = 0$, then $\phi''(c_i) = 2\alpha_3/(\alpha_4 + \alpha_5 c_i)^3 \geq 0$.

Outline

k-means clustering

$$\underset{c}{\text{minimize}} \quad \sum_{i=1}^{n} \sum_{j=1}^{k} c_{ij} \|x_i - \mu_j\|_2^2$$

Subject to the constraint that exactly one c_{ij} is 1 for every i.

1-means clustering

$$\underset{c}{\text{minimize}} \quad \sum_{i} w_i \left(c_i \|x_i - \mu\|_2^2 + (1 - c_i)\lambda_i \right)$$

weighted k-means clustering

$$\underset{c}{\text{minimize}} \quad \sum_{i=1}^{n} \sum_{j=1}^{k} w_i c_{ij} \|x_i - \mu_j\|_2^2$$

Subject to the constraint that exactly one c_{ij} is 1 for every i.

1-means clustering

$$\underset{c}{\text{minimize}} \quad \sum_{i} w_i \left(c_i \|x_i - \mu\|_2^2 + (1 - c_i)\lambda_i \right)$$

k-means clustering

weighted k-means clustering

$$\underset{c}{\text{minimize}} \quad \sum_{i=1}^{n} \sum_{j=1}^{k} w_i c_{ij} \|x_i - \mu_j\|_2^2$$

Subject to the constraint that exactly one c_{ij} is 1 for every i.

1-means clustering

$$\underset{c}{\text{minimize}} \quad \sum_{i} w_i \left(c_i \|x_i - \mu\|_2^2 + (1 - c_i)\lambda_i \right)$$

k-means clustering (cont.)

Optimal μ

Fix cluster assignment c_i, then

$$\mu = \frac{\sum_i w_i c_i x_i}{\sum_i w_i c_i}.$$

Optimal c

Fix μ, then c_i is 1 if $\|x_i - \mu\| < \lambda_i$, and 0 otherwise.

Radboud University

Kernels

$K(i,j) = \langle x_i, x_j \rangle$
so $\|x_i - x_j\|_2^2 = K(i,i) + K(j,j) - 2K(i,j)$.

Implicit centroid

The centroid is then a linear combination of points, $\mu = \sum_i \mu_i x_i$, giving

$$\|x_i - \mu\|_2^2 = K(i,i) - 2\sum_j \mu_j K(i,j) + \sum_{j,k} \mu_j K(j,k)\mu_k.$$

Optimal μ becomes

$$\mu_i = \frac{w_i c_i}{\sum_j w_j c_j}.$$

Kernel k-means clustering

Kernels

$K(i,j) = \langle x_i, x_j \rangle$
so $\|x_i - x_j\|_2^2 = K(i,i) + K(j,j) - 2K(i,j)$.

Implicit centroid

The centroid is then a linear combination of points, $\mu = \sum_i \mu_i x_i$, giving

$$\|x_i - \mu\|_2^2 = K(i,i) - 2\sum_j \mu_j K(i,j) + \sum_{j,k} \mu_j K(j,k) \mu_k.$$

Optimal μ becomes

$$\mu_i = \frac{w_i c_i}{\sum_j w_j c_j}.$$

Kernel k-means clustering (cont.)

Implicit centroid

$$\|x_i - \mu\|_2^2 = K(i,i) - 2\sum_j \mu_j K(i,j) + \sum_{j,k} \mu_j K(j,k) \mu_k.$$

$$\mu_i = \frac{w_i c_i}{\sum_j w_j c_j}.$$

1-means objective

$$\underset{c}{\text{minimize}} \quad \sum_i w_i \left(c_i \|x_i - \mu\|_2^2 + (1 - c_i)\lambda_i \right)$$

Kernel k-means clustering (cont.)

Radboud University

Implicit centroid

$$\|x_i - \mu\|_2^2 = K(i,i) - 2\sum_j \mu_j K(i,j) + \sum_{j,k} \mu_j K(j,k)\mu_k.$$

$$\mu_i = \frac{w_i c_i}{\sum_j w_j c_j}.$$

1-means objective

$$\underset{c}{\text{minimize}} \quad \sum_i \Big(w_i c_i K(i,i) - 2 w_i c_i \sum_j \mu_j K(i,j) + w_i c_i \sum_{j,k} \mu_j K(j,k)\mu_k + w_i(1 - c_i)\lambda_i\Big)$$

Kernel k-means clustering (cont.)

Radboud University

Implicit centroid

$$\|x_i - \mu\|_2^2 = K(i,i) - 2\sum_j \mu_j K(i,j) + \sum_{j,k} \mu_j K(j,k)\mu_k.$$

$$\mu_i = \frac{w_i c_i}{\sum_j w_j c_j}.$$

1-means objective

$$\begin{aligned} \underset{c}{\text{minimize}} \quad & \sum_i w_i c_i (K(i,i) - \lambda_i) + \sum_i w_i \lambda_i \\ & - \frac{\sum_{i,j} w_i c_i w_j c_j K(i,j)}{\sum_i w_i c_i}. \end{aligned}$$

Kernel k-means clustering (cont.)

Implicit centroid

$$\|x_i - \mu\|_2^2 = K(i,i) - 2\sum_j \mu_j K(i,j) + \sum_{j,k} \mu_j K(j,k)\mu_k.$$

$$\mu_i = \frac{w_i c_i}{\sum_j w_j c_j}.$$

1-means objective

$$\underset{c}{\text{minimize}} \quad 1 - \frac{\sum_{i,j} w_i c_i w_j c_j K(i,j)}{\sum_i w_i c_i},$$

taking $\lambda_i = K(i,i)$.

What is the kernel?

Idea

$$K = W^{-1} A W^{-1},$$

$$w_i = \sum_j a_{ij}$$

turns the objective into

$$\underset{c}{\text{minimize}} \quad 1 - \frac{\sum_{i,j} c_i c_j a_{ij}}{\sum_{i,j} c_i a_{ij}} = \phi(c),$$

We get conductance!

But this kernel is not positive definite.

What is the kernel?

Idea

$$K = W^{-1} A W^{-1},$$

$$w_i = \sum_j a_{ij}$$

turns the objective into

$$\underset{c}{\text{minimize}} \quad 1 - \frac{\sum_{i,j} c_i c_j a_{ij}}{\sum_{i,j} c_i a_{ij}} = \phi(c),$$

We get conductance!

But this kernel is not positive definite.

What is the kernel?

Idea

$$K = W^{-1} A W^{-1},$$

$$w_i = \sum_j a_{ij}$$

turns the objective into

$$\underset{c}{\text{minimize}} \quad 1 - \frac{\sum_{i,j} c_i c_j a_{ij}}{\sum_{i,j} c_i a_{ij}} = \phi(c),$$

We get conductance!

But this kernel is not positive definite.

Positive definite kernel

Radboud University

Add a diagonal

$$K = W^{-1}AW^{-1} + \sigma W^{-1}$$

The objective becomes

$$\underset{c}{\text{minimize}} \quad 1 - \frac{\sum_{i,j} c_i c_j a_{ij}}{\sum_{ij} c_i a_{ij}} - \sigma \frac{\sum_{i,j} c_i^2 a_{ij}}{\sum_{ij} c_i a_{ij}} = \phi_\sigma(c).$$

When $c_i \in \{0, 1\}$, $c_i^2 = c_i$, so the last term is constant.

Add a diagonal

$$K = W^{-1}AW^{-1} + \sigma W^{-1}$$

The objective becomes

$$\underset{c}{\text{minimize}} \quad 1 - \frac{\sum_{i,j} c_i c_j a_{ij}}{\sum_{ij} c_i a_{ij}} - \sigma \frac{\sum_{i,j} c_i^2 a_{ij}}{\sum_{ij} c_i a_{ij}} = \phi_\sigma(c).$$

When $c_i \in \{0,1\}$, $c_i^2 = c_i$, so the last term is constant.

A look at local optima

Relaxing the optimization problem

$$\begin{aligned} \text{minimize} \quad & \phi_\sigma(c) \\ \text{subject to} \quad & 0 \le c_i \le 1 \quad \text{for all } i \in V. \end{aligned} \tag{1}$$

Theorem

When $\sigma \geq 2$, every discrete community c is a local optimum of (1).

In practice

Higher $\sigma \Rightarrow$ more clusters are local optima.

A look at local optima

Relaxing the optimization problem

$$\begin{aligned} \text{minimize} \quad & \phi_\sigma(c) \\ \text{subject to} \quad & 0 \le c_i \le 1 \quad \text{for all } i \in V. \end{aligned} \tag{1}$$

Theorem

When $\sigma \geq 2$, every discrete community c is a local optimum of (1).

In practice

Higher $\sigma \Rightarrow$ more clusters are local optima.

Outline

The problem

Given some *seed nodes* s find the cluster c that contains s.

As constrained optimization

$$\begin{aligned} &\text{minimize} && \phi_\sigma(c) && \\ &\text{subject to} && 0 \leq c_i \leq 1 && \text{for all } i \in V. \\ &\text{and} && c_i = 1 && \text{for all } i \in s \end{aligned}$$

The problem

Given some *seed nodes* s find the cluster c that contains s.

As constrained optimization

$$\begin{aligned} \text{minimize} \quad & \phi_\sigma(c) \\ \text{subject to} \quad & 0 \le c_i \le 1 \quad && \text{for all } i \in V. \\ \text{and} \quad & c_i = 1 && \text{for all } i \in s \end{aligned}$$

Projected Gradient Descent

Radboud University

Gradient descent

$$c^{(0)} = s,$$
$$c^{(t+1)} = p(c^{(t)} - \alpha^{(t)} \nabla \phi_\sigma(c^{(t)})).$$

Project onto valid set

$$p(c) = \operatorname*{argmin}_{c',\ \text{s.t.}\ 0 \le c'_i \le 1, c'_i \ge s_i} \|c - c'\|_2^2,$$

This is simply

$$p(c) = \max(s, \min(1, c)).$$

Projected Gradient Descent

Gradient descent

$$c^{(0)} = s,$$
$$c^{(t+1)} = p(c^{(t)} - \alpha^{(t)} \nabla \phi_\sigma(c^{(t)})).$$

Project onto valid set

$$p(c) = \operatorname*{argmin}_{c',\ \text{s.t.}\ 0 \leq c'_i \leq 1, c'_i \geq s_i} \|c - c'\|_2^2,$$

This is simply

$$p(c) = \max(s, \min(1, c)).$$

Projected Gradient Descent

Gradient descent

$$c^{(0)} = s,$$
$$c^{(t+1)} = p(c^{(t)} - \alpha^{(t)} \nabla \phi_\sigma (c^{(t)})).$$

Project onto valid set

$$p(c) = \underset{c',\ \text{s.t. } 0 \le c'_i \le 1, c'_i \ge s_i}{\operatorname{argmin}} \|c - c'\|_2^2,$$

This is simply

$$p(c) = \max(s, \min(1, c)).$$

Expectation Maximization

Radboud University

E-step:

$$c_i \leftarrow 1 \text{ if } \|x_i - \mu\| \leq \lambda_i, \text{ and 0 otherwise.}$$

M-step:

$$\mu_i \leftarrow \frac{w_i c_i}{\sum_j w_j c_j}.$$

Together

If you work this out, this is equivalent to

$$c^{(0)} = s,$$
$$c^{(t+1)} = \{i \mid \nabla\phi_\sigma(c^{(t)}) < 0\} \cup s.$$

This is gradient descent with infinite step size.

Expectation Maximization

Radboud University

E-step:

$$c_i \leftarrow 1 \text{ if } \|x_i - \mu\| \leq \lambda_i, \text{ and 0 otherwise.}$$

M-step:

$$\mu_i \leftarrow \frac{w_i c_i}{\sum_j w_j c_j}.$$

Together

If you work this out, this is equivalent to

$$c^{(0)} = s,$$
$$c^{(t+1)} = \{i \mid \nabla\phi_\sigma(c^{(t)}) < 0\} \cup s.$$

This is gradient descent with infinite step size.

Expectation Maximization

E-step:

$$c_i \leftarrow 1 \text{ if } \|x_i - \mu\| \leq \lambda_i, \text{ and 0 otherwise.}$$

M-step:

$$\mu_i \leftarrow \frac{w_i c_i}{\sum_j w_j c_j}.$$

Together

If you work this out, this is equivalent to

$$\begin{aligned} c^{(0)} &= s, \\ c^{(t+1)} &= \{i \mid \nabla\phi_\sigma(c^{(t)}) < 0\} \cup s. \end{aligned}$$

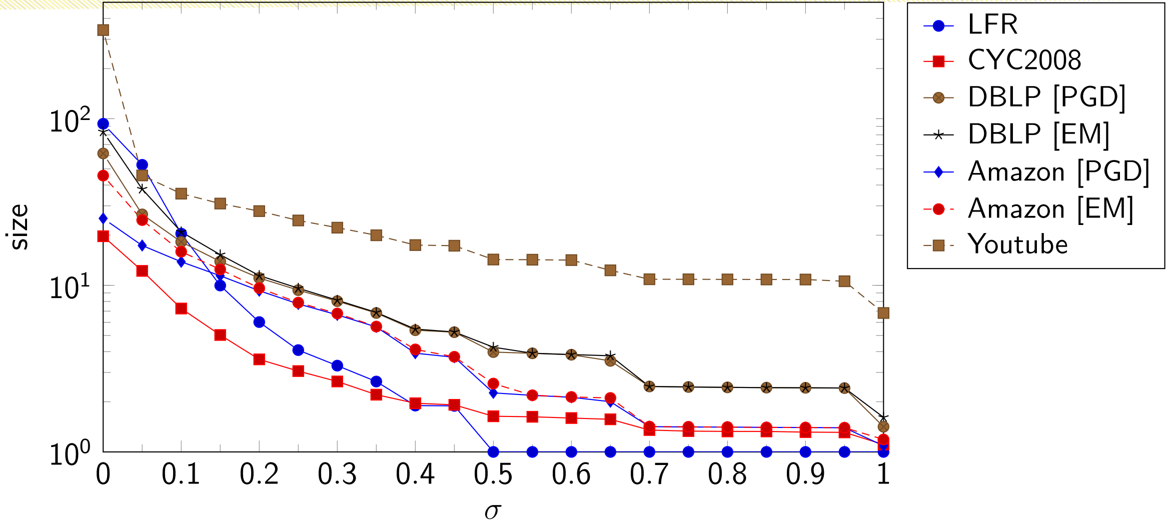
This is gradient descent with infinite step size.

Outline

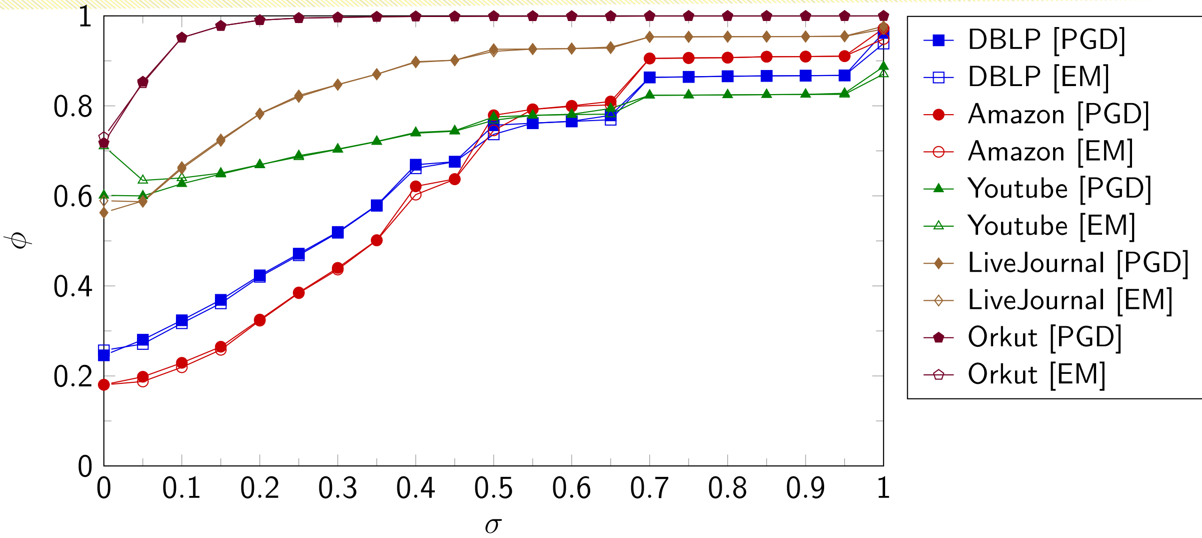
Datasets

Dataset	#node	#edge	clus.c.	#comm
LFR (om=2)	5000	25123	0.021	146
CYC2008	6230	6531	0.121	408
Amazon	334863	925872	0.079	151037
DBLP	317080	1049866	0.128	13477
Youtube	1134890	2987624	0.002	8385
LiveJournal	3997962	34681189	0.045	287512
Orkut	3072441	117185083	0.014	6288363

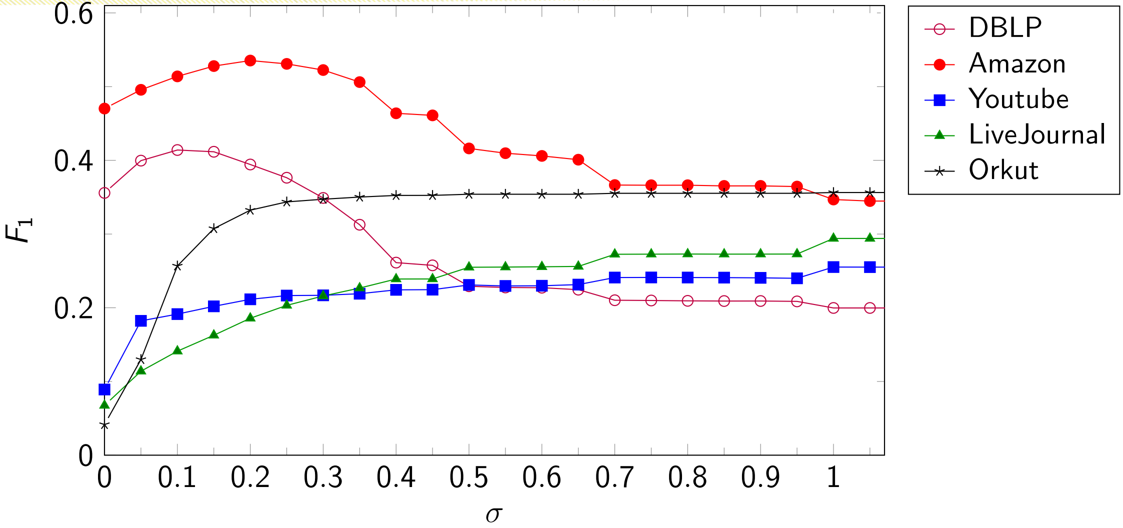
Cluster size



Conductance



Ground truth recovery



Choice of σ

- Choice of σ is important.
- Heuristic: Pick σ to maximize community density.

F_1 scores

Dataset	PGDc-0	PGDc-d	EMc-0	EMc-d	YL	HK	PPR
LFR (om=1)	**0.967**	0.185	0.868	0.187	0.203	0.040	0.041
LFR (om=2)	**0.483**	0.095	0.293	0.092	0.122	0.039	0.041
LFR (om=3)	**0.275**	0.085	0.158	0.083	0.110	0.037	0.039
LFR (om=4)	**0.178**	0.074	0.100	0.072	0.092	0.032	0.034
Karate	0.831	0.472	0.816	0.467	0.600	0.811	**0.914**
Football	**0.792**	**0.816**	0.766	**0.805**	**0.816**	0.471	0.283
Pol.Blogs	**0.646**	0.141	**0.661**	0.149	0.017	**0.661**	0.535
Pol.Books	0.596	0.187	0.622	0.197	0.225	**0.641**	**0.663**
Flickr	0.098	0.027	0.097	0.027	0.013	0.054	**0.118**
CYC	0.474	**0.543**	0.455	**0.543**	**0.526**	0.336	0.294
Amazon	0.470	**0.522**	0.425	**0.522**	0.493	0.245	0.130
DBLP	**0.356**	**0.369**	0.317	**0.371**	0.341	0.214	0.210
Youtube	0.089	**0.251**	0.073	**0.248**	0.228	0.037	0.071
LiveJournal	0.067	**0.262**	0.059	**0.259**	0.183	0.035	0.049
Orkut	0.042	**0.231**	0.033	**0.231**	0.171	0.057	0.033

Outline

Radbound University

Network community detection with conductance

The relation to k-means clustering

Algorithms

Experiments

Conclusions

Conclusions

- Conductance is related to 1-means clustering.
- Similarly, EM is the limiting case of Projected Gradient Descent.
- High conductance clusters often correspond to 'true' clusters.
- ...but they can be very large.

1-means clustering and conductance

Twan van Laarhoven

Radboud University Nijmegen
Open Universiteit Nederland

November 27th, 2017