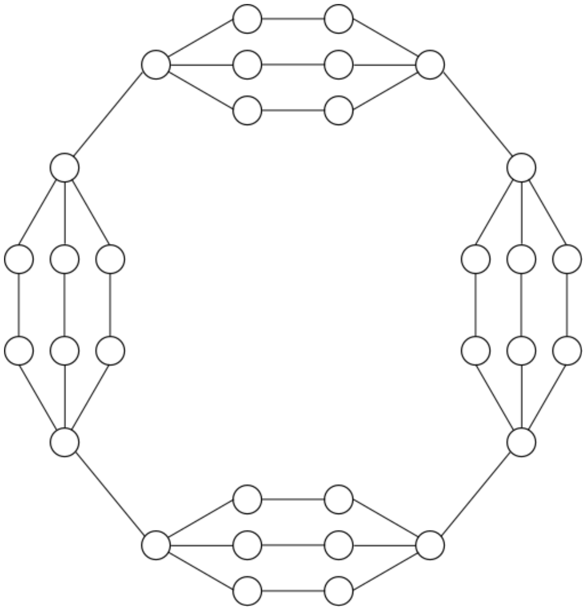
Graph-based semi-supervised learning for complex networks

Leto Peel

Université catholique de Louvain

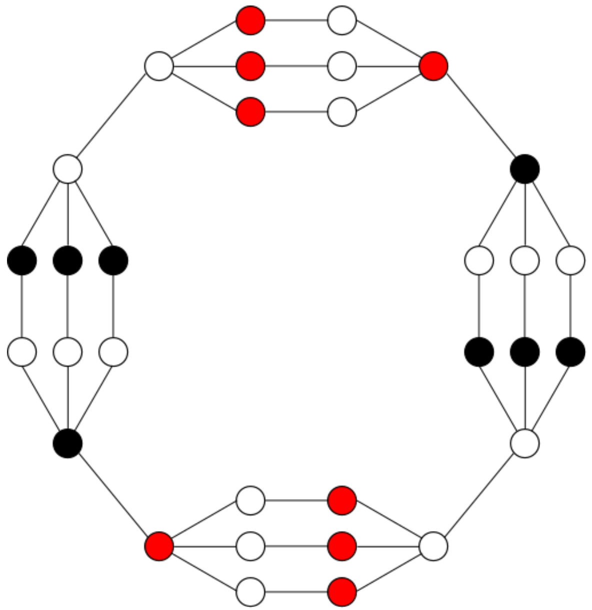
@PiratePeel

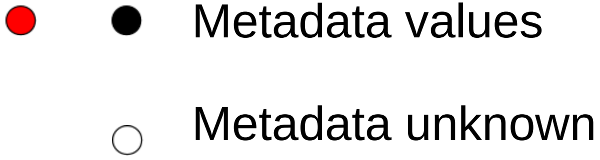
Here is a network



social networks
food webs
internet
protein interactions

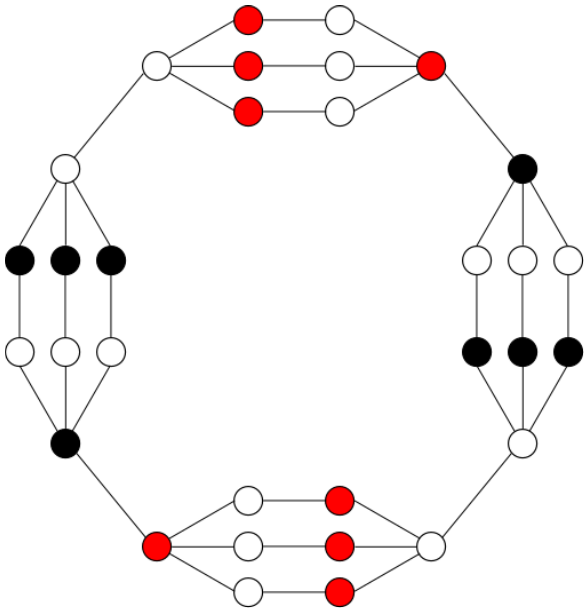
Network nodes can have properties or attributes (metadata)





social networks *age, sex, ethnicity, race, etc.*
food webs *feeding mode, species body mass, etc.*
internet *data capacity, physical location, etc.*
protein interactions *molecular weight, association with cancer, etc.*

Network nodes can have properties or attributes (metadata)



Metadata values

Metadata unknown

Can we predict the unknown metadata values?

social networks *age, sex, ethnicity, race, etc.*
food webs *feeding mode, species body mass, etc.*
internet *data capacity, physical location, etc.*
protein interactions *molecular weight, association with cancer, etc.*

Now, let's talk about supervised learning...

input $\qquad\qquad$ output

$$X \xrightarrow{f} Y$$

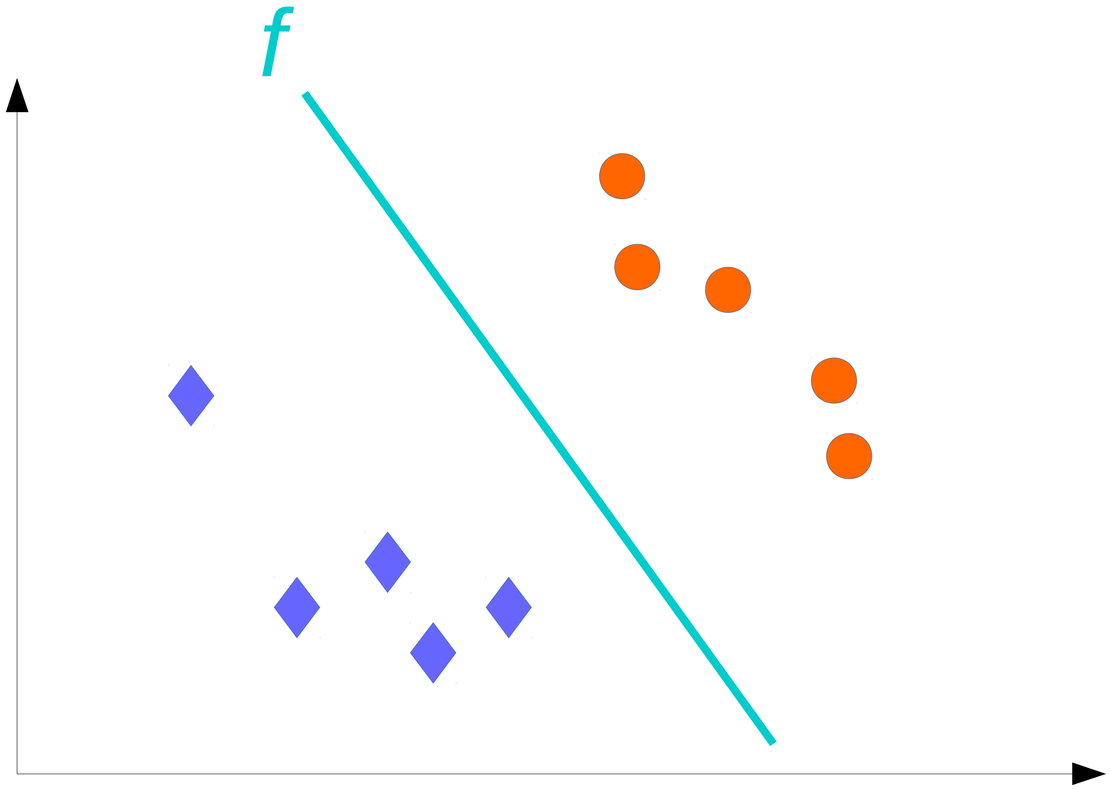
feature vector $\qquad$ *discrete label*

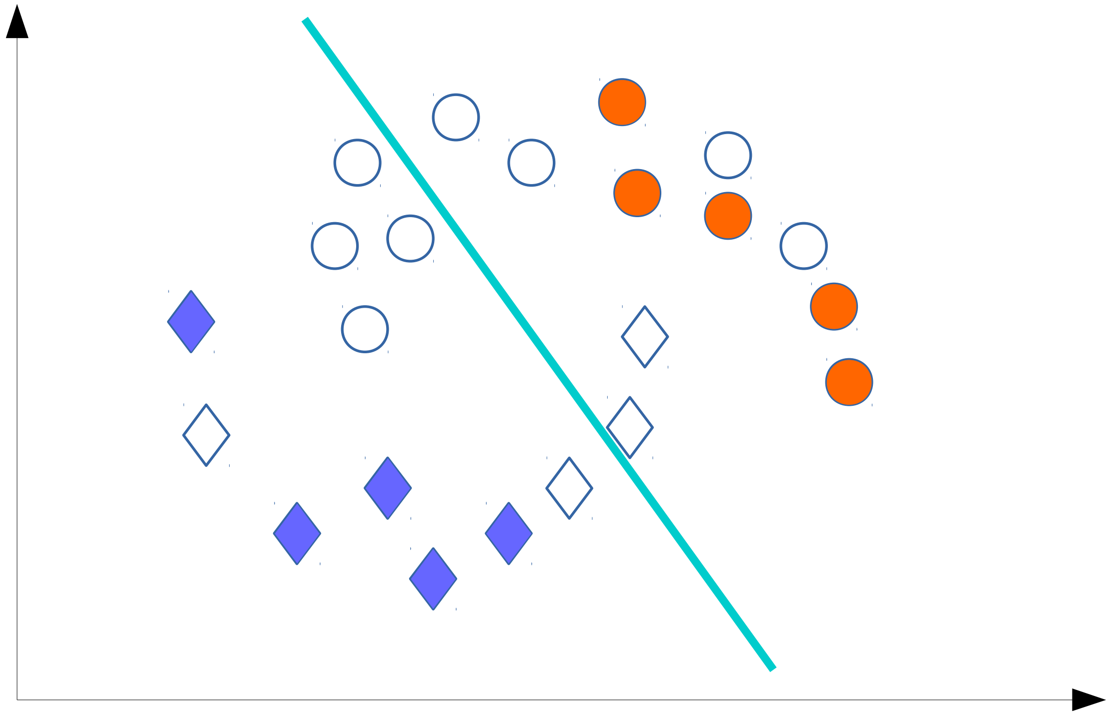
Training

$$\{(X,Y)\}_{train} \xrightarrow{\textit{inference}} f$$

Predict

$$f(X_{test}) \xrightarrow{\textit{classification}} \tilde{Y}$$


f



Now, let's talk about *semi*-supervised learning...

input $X \xrightarrow{f} Y$ output

feature vector — *discrete label*

Training

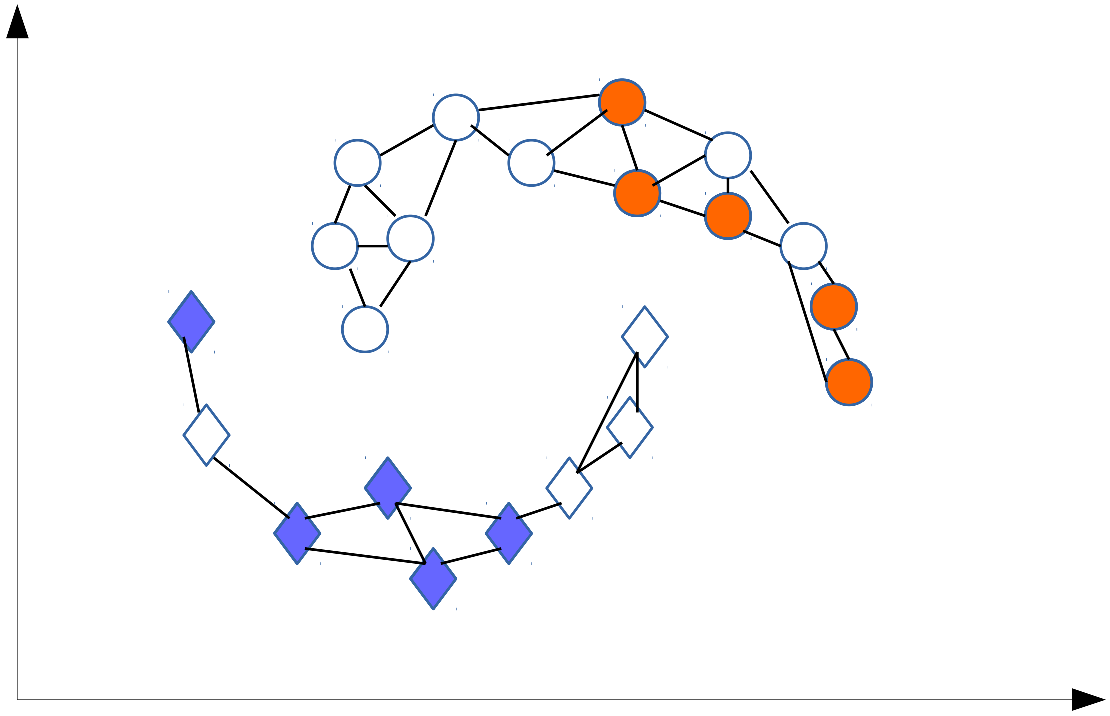
$$\{(X,Y)\}_{train} \xrightarrow{\textit{inference}} f$$

use all available data for training the classifier

Predict

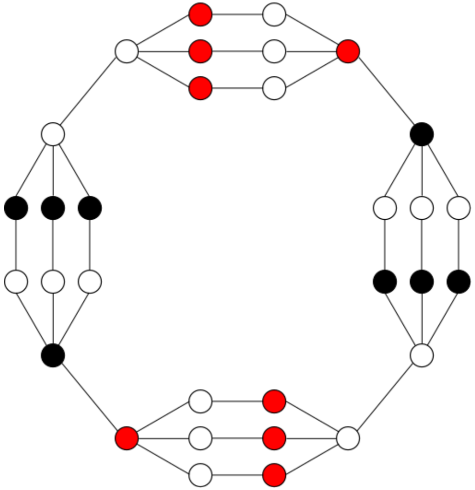
$$f(X_{test}) \xrightarrow{\textit{classification}} \tilde{Y}$$

Graph-based semi-supervised learning



Construct a graph based on similarity in X and propagate label information around the graph

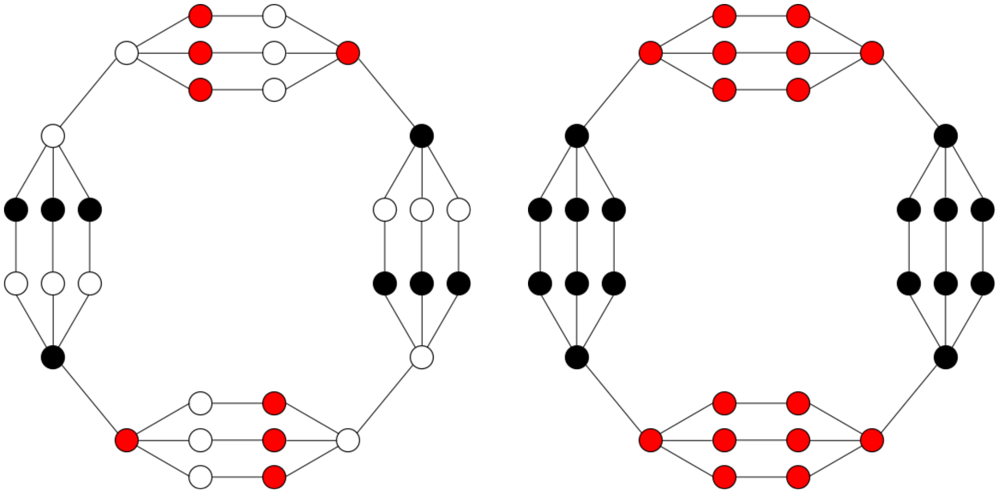
Semi-supervised learning in complex networks



Metadata values

Metadata unknown

Semi-supervised learning in complex networks

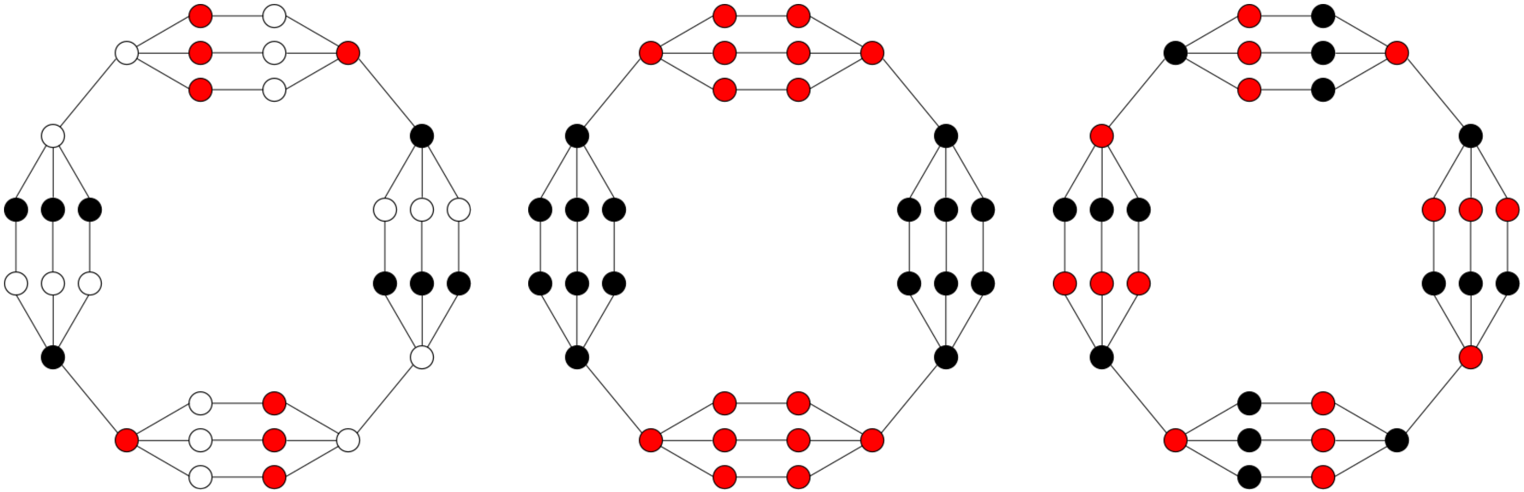


assortative

Metadata values

Metadata unknown

Semi-supervised learning in complex networks

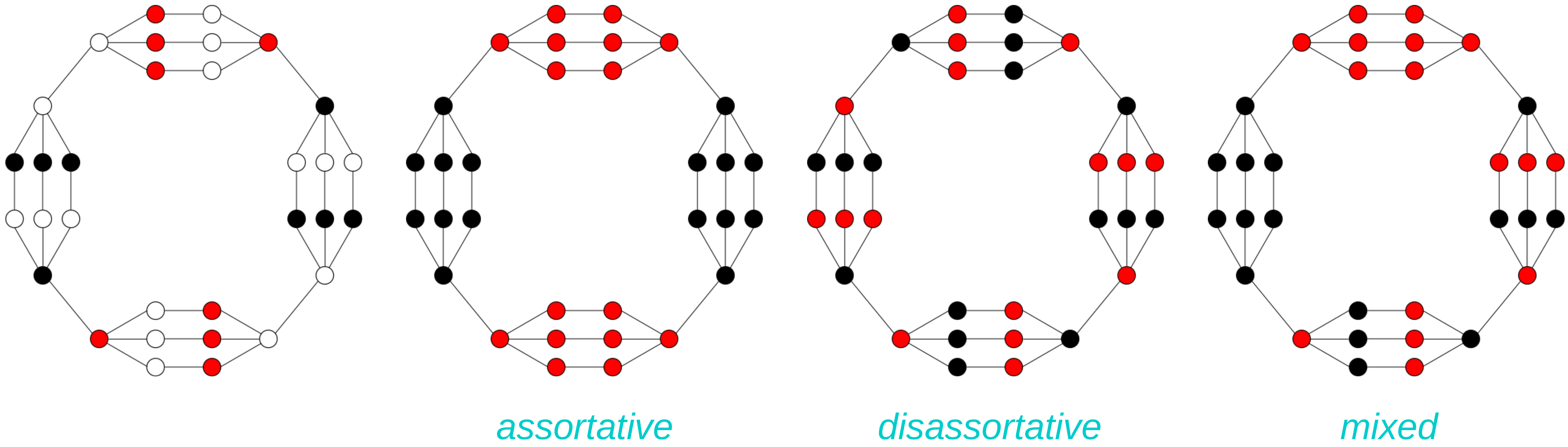


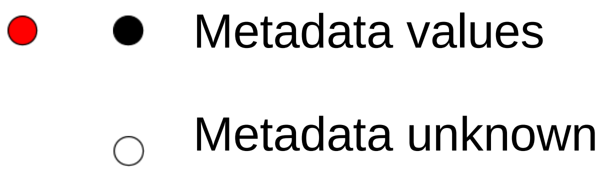
assortative *disassortative*

Metadata values

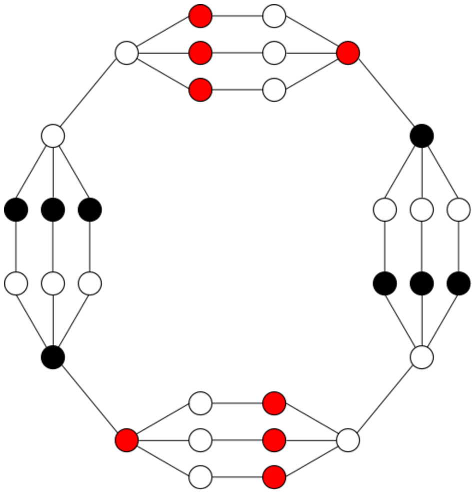
Metadata unknown

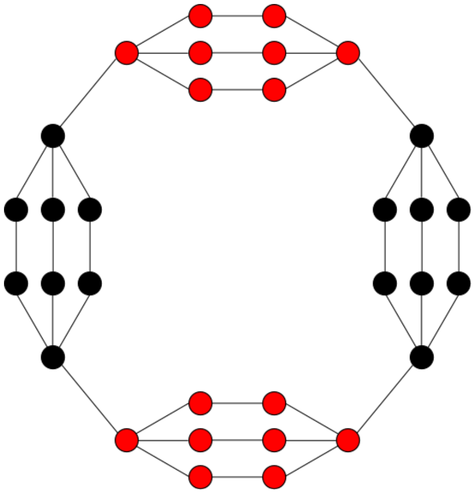
Semi-supervised learning in complex networks



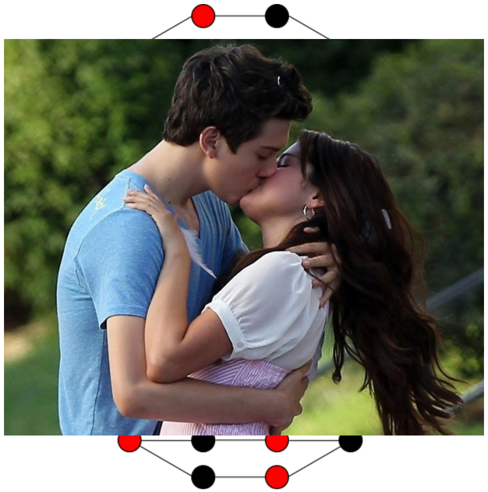


Semi-supervised learning in relational networks

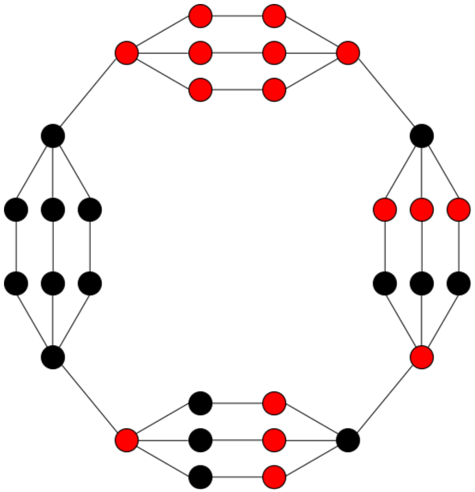




assortative



disassortative

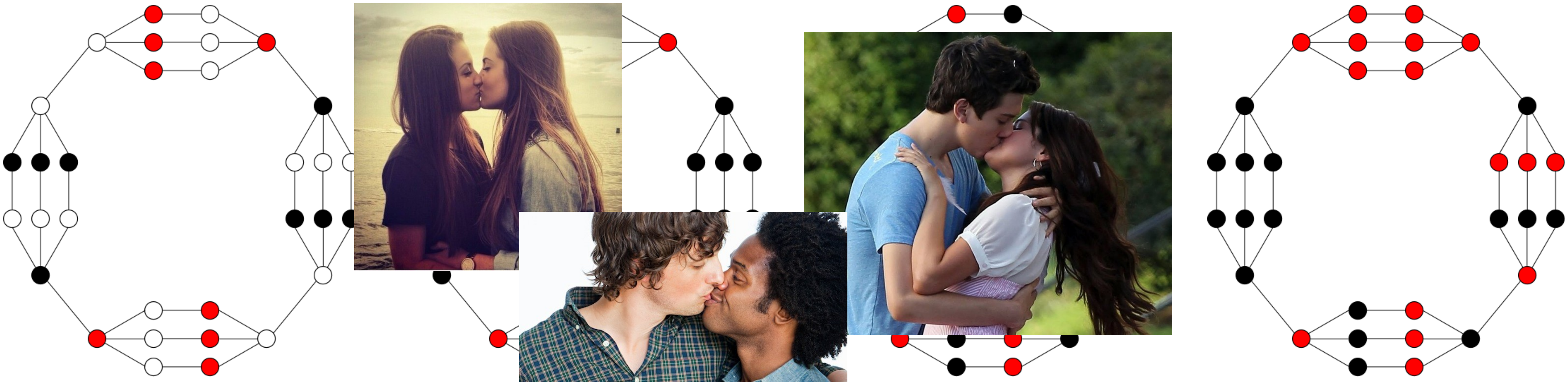


mixed

Metadata values

Metadata unknown

Semi-supervised learning in relational networks

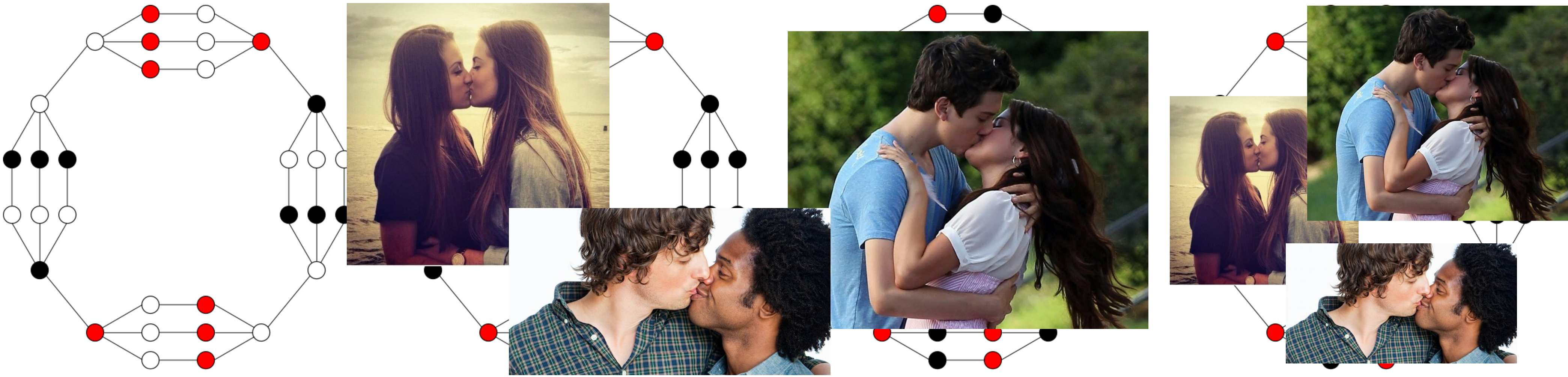


assortative *disassortative* *mixed*

Metadata values

Metadata unknown

Semi-supervised learning in relational networks



assortative *disassortative* *mixed*

Metadata values

Metadata unknown

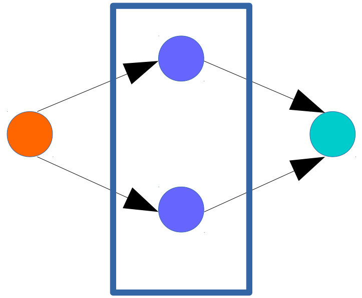
Naive application of label propagation does not work if we don't know how classes interact

Naive application of label propagation does not work if we don't know how classes interact

Solution: Construct a similarity graph based on the relational network

DEFINITION 3. *Structural equivalence [24]* — *If* $\equiv$ *is an equivalence relation on* $\mathcal{V}_r$ *then* $\equiv$ *is a structural equivalence iff* $\forall a, b, c \in \mathcal{V}_r$ *and* $a \equiv b$*:*

1. $\forall (a \rightarrow c) \in \mathcal{E}_r$, *then* $(b \rightarrow c) \in \mathcal{E}_r$;
2. $\forall (a \leftarrow c) \in \mathcal{E}_r$, *then* $(b \leftarrow c) \in \mathcal{E}_r$.



Structurally equivalent nodes

Lorrain & White, Structural equivalence of individuals in social networks. J. Math. Sociol., 1971

Common neighbours

cosine similarity is a measure of how structurally equivalent two nodes are

$$S_{a,b} = \frac{\sum_c A_{ac} A_{cb}}{\sqrt{\sum_c A_{ac}^2} \sqrt{\sum_c A_{cb}^2}}$$

$$\mathbf{S}_{\leftrightarrow} = \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$$

cosine label propagation

$$\mathbf{F}_{t+1} = \mathbf{Z}^{-1} \left((1-\alpha) \mathbf{B} + \alpha \mathbf{S}_{\leftrightarrow} \mathbf{F}_t \right)$$

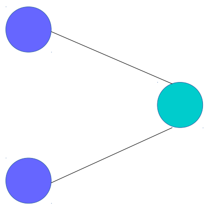
Neighbours of neighbours

the set of neighbours of a node's neighbours contain all structurally equivalent nodes

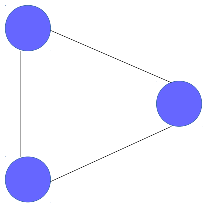
two-step label propagation

$$\mathbf{F}_{t+1} = \mathbf{Z}^{-1}((1-\alpha)\mathbf{B} + \alpha(\mathbf{LL})^{\beta}\mathbf{F}_t)$$

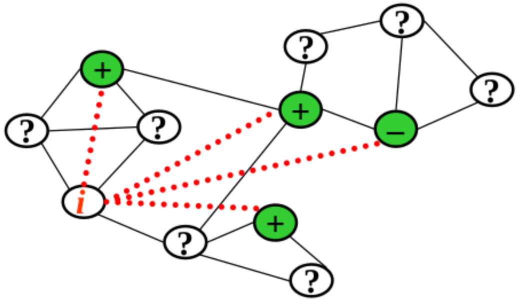
Why are paths of length 2 important?



bipartite / diassortative
negative auto-correlation



presence of triangles
in assortative relations

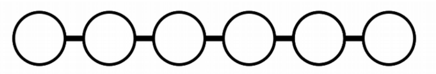


Gallagher et al. Using ghost edges for classification in sparsely labeled networks, KDD 2008

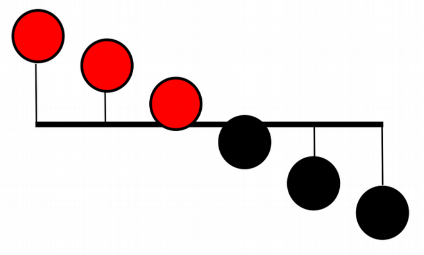
Why are paths of length 2 important?

Label propagation is an eigenvector problem

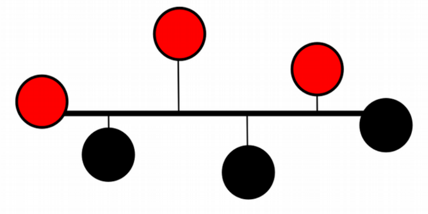
$$\mathbf{L} = \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$$ has eigenvalues in [-1,1]



most positive



most negative



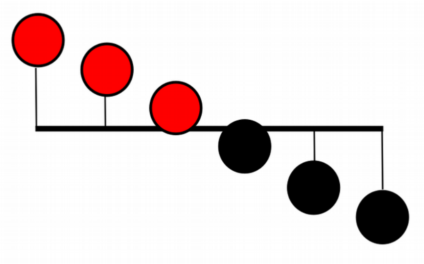
Why are paths of length 2 important?

Label propagation is an eigenvector problem

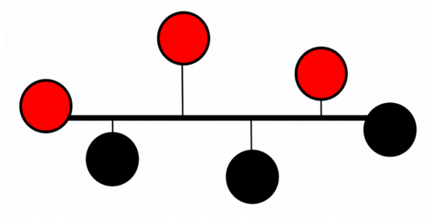
$$\mathbf{L} = \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$$ has eigenvalues in [-1,1]

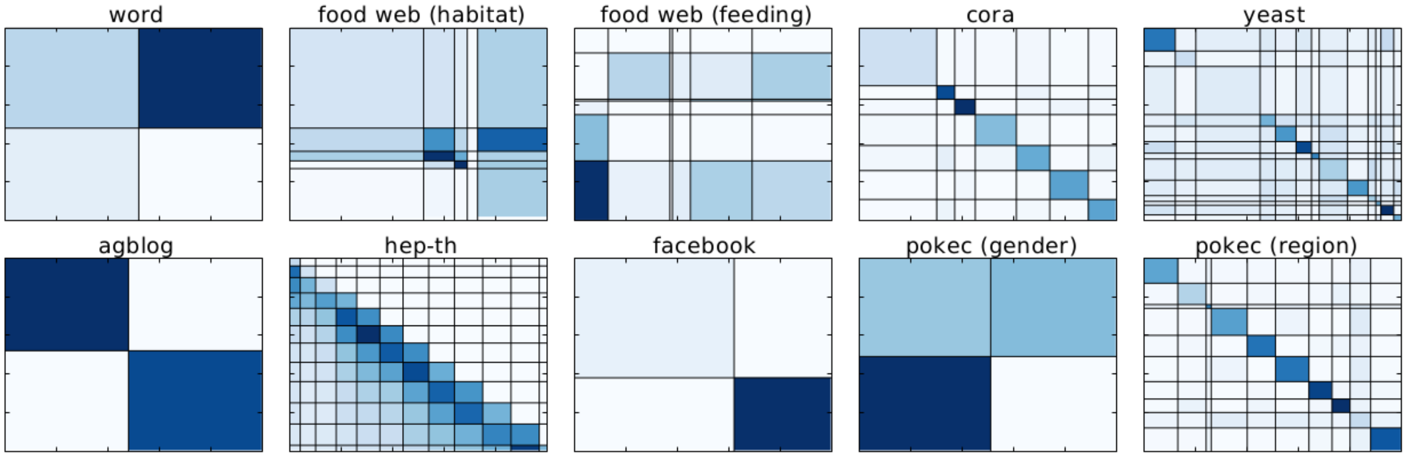
When we consider even path lengths using L^2 (or A^2 in the case of cosine LP) the eigenvectors remain unchanged, but the eigenvalues are all positive

positive



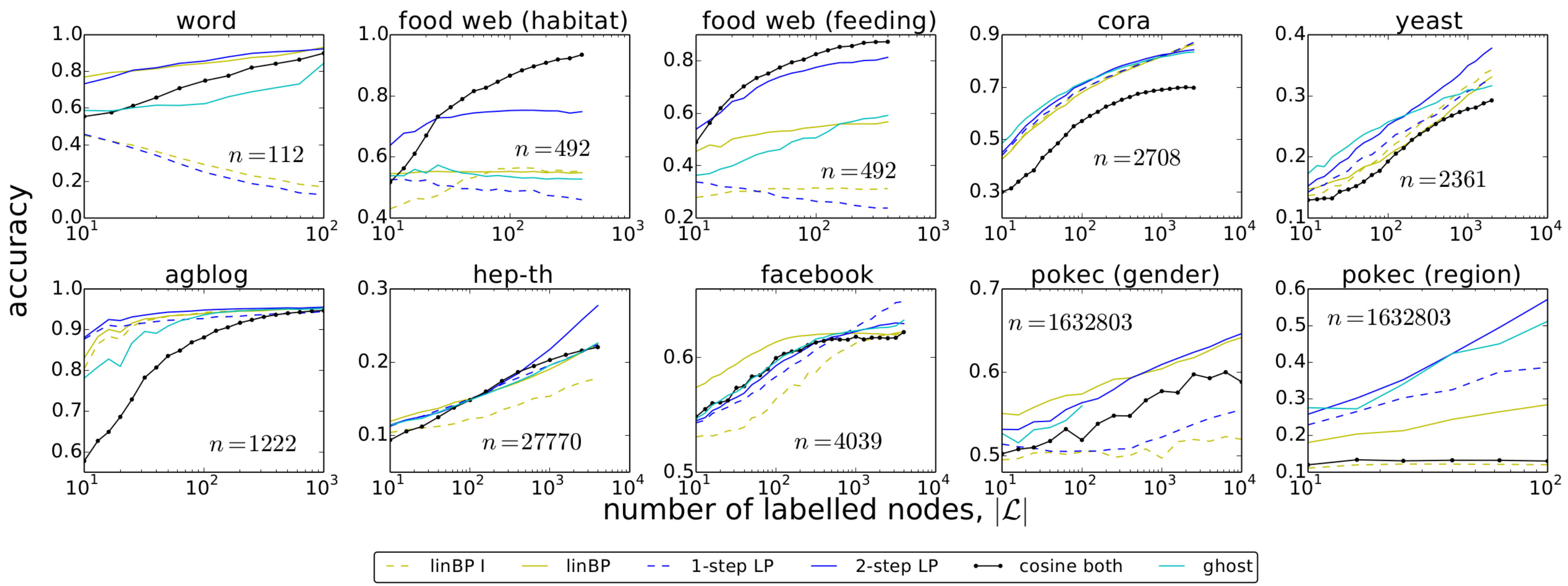
positive





Network (label)	n	$\tilde{n}$	m	ℓ
word (adj/noun) [27]	112	0	569	2
foodweb (habitat) [7]	492	4	16330	5
foodweb (feeding) [7]	492	4	16330	6
cora (subject) [35]	2708	0	5429	7
yeast (function) [8]	2361	0	7182	13
agblog (political) [1]	1222	0	33428	2
hep-th (year) [15]	27770	0	352807	12
facebook (gender) [25]	4039	0	176468	2
pokec (gender) [37]	1632803	163	30622564	2
pokec (region) [37]	1632803	163	30622564	10

Gratuitous Comp. Sci. "My curve is better than your curve" slide



Take home messages…

1) Complex networks are not (necessarily) the same as similarity graphs
    - we should adapt our methods accordingly

Take home messages...

1) Complex networks are not (necessarily) the same as similarity graphs
    - we should adapt our methods accordingly

2) Machine Learning for Complex Networks does not require representing nodes as feature vectors
    - use Network Science!

Advertisement

The Winter Workshop on Complex Systems is a one-week workshop where young researchers from all over the world gather together for discussing about complexity science and engaging into novel research projects.

Winter Workshop on Complex Systems 2019

Zakopane, Poland

February 2019

Applications now open!

http://wwcs2019.org/

February 4-8th 2019

Zakopane, Poland

For more information...

Peel, Graph-based semi-supervised learning for relational networks. SIAM International Conference on Data Mining, 2017

https://arxiv.org/abs/1612.05001

Contact:

leto.peel@uclouvain.be

@PiratePeel

regularisation parameter

predicted labels

Linear operator

known labels

$$\min_{\mathbf{F}} \quad \frac{1}{2}\mathrm{tr}(\mathbf{F}^{\mathbf{T}}\mathbf{L}\mathbf{F}) + \lambda\|\mathbf{F}-\mathbf{B}\|^2 \quad \text{s.t.} \quad \mathbf{F}\mathbb{1} = \mathbb{1}$$

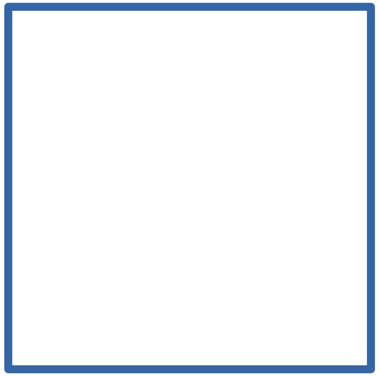
regularisation parameter

predicted labels

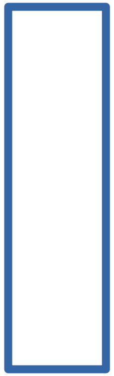
Linear operator

known labels

$$\min_{\mathbf{F}} \quad \frac{1}{2}\mathrm{tr}(\mathbf{F}^{\mathbf{T}}\mathbf{L}\mathbf{F}) + \lambda\|\mathbf{F} - \mathbf{B}\|^2 \quad \text{s.t.} \quad \mathbf{F}\mathbb{1} = \mathbb{1}$$

L = 

N x N
(graph connectivity)

B= 

N x C
1 (or 0) if we know node belongs to class (or not)
1/C otherwise

Initialise F=B

regularisation parameter

predicted labels

Linear operator

known labels

$$\min_{\mathbf{F}} \quad \frac{1}{2}\underbrace{\mathrm{tr}(\mathbf{F}^{\mathbf{T}}\mathbf{L}\mathbf{F})}_{\text{smoothness}} + \lambda \underbrace{\|\mathbf{F}-\mathbf{B}\|^2}_{\text{consistency}} \quad \text{s.t.} \quad \mathbf{F}\mathbb{1} = \mathbb{1}$$

predicted labels

known labels

$$\min_{\mathbf{F}} \quad \frac{1}{2}\mathrm{tr}(\mathbf{F}^{\mathbf{T}}\mathbf{L}\mathbf{F}) + \lambda\|\mathbf{F} - \mathbf{B}\|^2 \quad \text{s.t.} \quad \mathbf{F}\mathbb{1} = \mathbb{1}$$

$$\mathbf{L} = \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}} \text{ and } \lambda = (\frac{1}{\alpha} - 1)$$

not I – $D^{-(1/2)}AD^{-(1/2)}$ since we require the “smoothest” eigenvector to be dominant (associated with the largest eigenvalue)

Zhou et al. Learning with local and global consistency, NIPS 2003 predicted labels

known labels

$$\min_{\mathbf{F}} \quad \frac{1}{2}\mathrm{tr}(\mathbf{F}^{\mathbf{T}}\mathbf{L}\mathbf{F}) + \lambda \|\mathbf{F} - \mathbf{B}\|^2 \quad \text{s.t.} \quad \mathbf{F}\mathbb{1} = \mathbb{1}$$

$$\mathbf{L} = \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}} \text{ and } \lambda = \left(\frac{1}{\alpha} - 1\right)$$

Solve using the power method:

$$\mathbf{F}_{t+1} = \mathbf{Z}^{-1}((1-\alpha)\mathbf{B} + \alpha\mathbf{L}\mathbf{F}_t)$$

Zhou et al. Learning with local and global consistency, NIPS 2003