# Implementation Issues for Latent Space Embedding

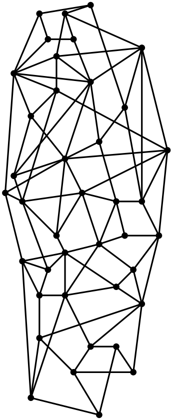
Minkyoung Cho, David Mount, and Eunhui Park

Department of Computer Science
University of Maryland, College Park

MURI Meeting – May 25, 2010

# Motivation

- Social networks are used to represent a variety of relational data.
- Social networks exhibit structural features:
  - Transitivity
  - Homophily on attributes
  - Clustering
- The likelihood of a tie is often correlated with the similarity of attributes of the actors. (E.g., geography, age, ethnicity, income).
- These attributes may be observed or unobserved.
- A subset of nodes with many ties between them may indicate clustering with respect to an underlying (latent) social space.



# Latent Space Embedding (LSE)

## Hypothesis

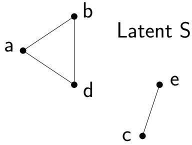
The likelihood of relational ties in social networks depends on the similarity of attributes in an unobserved latent space.

## Problem Statement

Given a network $Y = [y_{i,j}]$ with $n$ nodes, estimate a set of positions $Z = \{z_1, \ldots, z_n\}$ in $\mathbb{R}^d$ that best describes this network relative to some model.

Network

| | a | b | c | d | e |
|---|---|---|---|---|---|
| a | - | 1 | 0 | 1 | 0 |
| b | 1 | - | 0 | 1 | 0 |
| c | 0 | 0 | - | 0 | 1 |
| d | 1 | 1 | 0 | - | 0 |
| e | 0 | 0 | 1 | 0 | - |




Latent Space

# LSE — Stochastic Model

## Input

- $Y$: An $n \times n$ sociomatrix
  ($y_{i,j} = 1$ if there is a tie between $i$ and $j$)

## Model Parameters

- $Z$: The positions of $n$ individuals, $\{z_1, \ldots, z_n\}$ in latent space
- $\alpha$: Real-valued scaling parameter

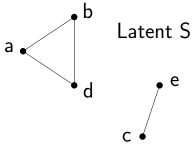
## Stochastic Model [HRH02]

Ties are statistically independent:

$$\Pr[Y \mid Z, \alpha] \triangleq \prod_{i \neq j} \Pr[y_{i,j} \mid z_i, z_j, \alpha]$$

Network

| | a | b | c | d | e |
|---|---|---|---|---|---|
| a | - | 1 | 0 | 1 | 0 |
| b | 1 | - | 0 | 1 | 0 |
| c | 0 | 0 | - | 0 | 1 |
| d | 1 | 1 | 0 | - | 0 |
| e | 0 | 0 | 1 | 0 | - |




Latent Space

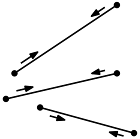
# LSE — Stochastic Model

## Logistic Regression Model [HRH02]

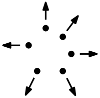
$$\log \text{odds}(y_{i,j} = 1 \mid z_i, z_j, \alpha) \;=\; \alpha - \|z_i - z_j\|.$$

Define $\eta_{i,j} \triangleq \alpha - \|z_i - z_j\|$. We have

$$\log \Pr[Y \mid \eta] \;=\; \sum_{i \neq j} (\eta_{i,j} y_{i,j} - \log(1 + e^{\eta_{i,j}})).$$



Stretch



Spread

To maximize $\Pr[Y \mid \eta]$:

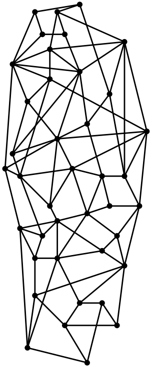
- Minimize Stretch: $\sum_{i \neq j} \eta_{i,j} y_{i,j} \;\Rightarrow$ Shrinks long edges.
- Maximize Spread: $-\sum_{i \neq j} \log(1 + e^{\eta_{i,j}}) \;\Rightarrow$ Keeps points apart.

# LSE — Efficient cost computation

## Computational Problem

Given an $n \times n$ matrix $Y$, determine $Z$ and $\alpha$ to maximize $\Pr[Y \mid Z, \alpha]$.

- Method: Markov-Chain Monte Carlo (MCMC):
  - Perturb current point locations: $Z \rightarrow Z^*$.
  - Compute change in probability: $\rho = \frac{\Pr[Y|Z^*,\alpha]}{\Pr[Y|Z,\alpha]}$.
  - Accept change with probability $\min(1, \rho)$.
  - Rinse and repeat.
- Issues:
  - Computing $\Pr[Y \mid Z, \alpha]$ takes quadratic time.
  - Use a spatial index to store spatial relationships.
  - The index must be dynamic.

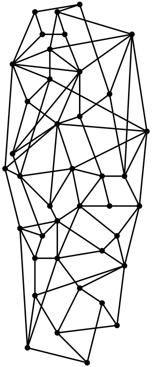


# LSE — Efficient cost computation

## Computational Problem

Given an $n \times n$ matrix $Y$, determine $Z$ and $\alpha$ to maximize $\Pr[Y \mid Z, \alpha]$.

- Method: Markov-Chain Monte Carlo (MCMC):
  - Perturb current point locations: $Z \rightarrow Z^*$.
  - Compute change in probability: $\rho = \frac{\Pr[Y|Z^*,\alpha]}{\Pr[Y|Z,\alpha]}$.
  - Accept change with probability $\min(1, \rho)$.
  - Rinse and repeat.
- Issues:
  - Computing $\Pr[Y \mid Z, \alpha]$ takes quadratic time.
  - Use a spatial index to store spatial relationships.
  - The index must be dynamic.

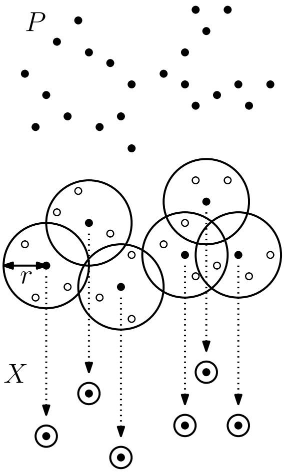


# Computational Tools – Nets

## Net

$P$ is a finite set of points in a $\mathbb{R}^d$. Given $r > 0$, an $r$-net for $P$ is a subset $X \subseteq P$ such that,

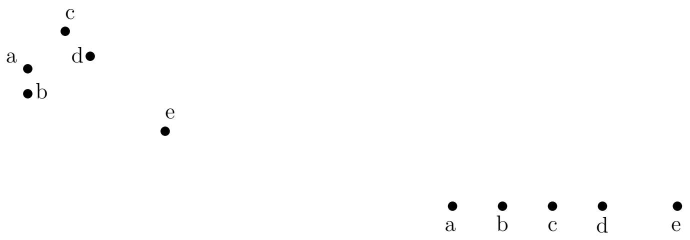
$$\max_{p \in M} dist(p, X) < r \quad \text{and}$$
$$\min_{\substack{x, x' \in X \\ x \neq x'}} \|x - x'\| \geq r.$$



# Net Trees

## Net Tree

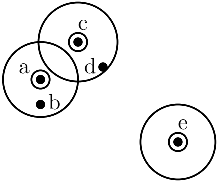
- The leaves of the tree consists of the points of $P$.
- The tree is based on a series of nets, $P^{(1)}, P^{(2)}, \ldots, P^{(h)}$, where $P^{(i)}$ is a $(2^i)$-net for $P^{(i-1)}$.
- Each node on level $i-1$ is associated with a parent, at level $i$, which lies lies within distance $2^i$.

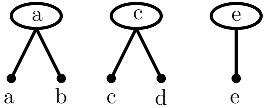


# Net Trees

## Net Tree

- The leaves of the tree consists of the points of $P$.
- The tree is based on a series of nets, $P^{(1)}, P^{(2)}, \ldots, P^{(h)}$, where $P^{(i)}$ is a $(2^i)$-net for $P^{(i-1)}$.
- Each node on level $i-1$ is associated with a parent, at level $i$, which lies lies within distance $2^i$.

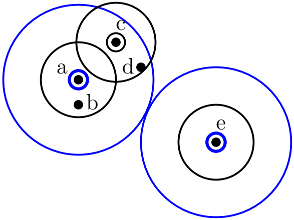


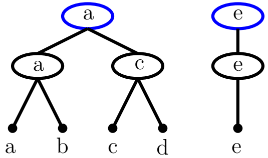



# Net Trees

## Net Tree

- The leaves of the tree consists of the points of $P$.
- The tree is based on a series of nets, $P^{(1)}, P^{(2)}, \ldots, P^{(h)}$, where $P^{(i)}$ is a $(2^i)$-net for $P^{(i-1)}$.
- Each node on level $i-1$ is associated with a parent, at level $i$, which lies lies within distance $2^i$.

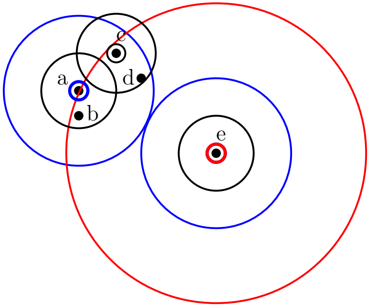


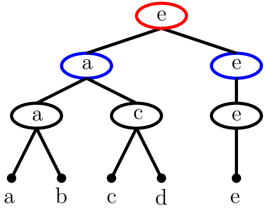



# Net Trees

## Net Tree

- The leaves of the tree consists of the points of $P$.
- The tree is based on a series of nets, $P^{(1)}, P^{(2)}, \ldots, P^{(h)}$, where $P^{(i)}$ is a $(2^i)$-net for $P^{(i-1)}$.
- Each node on level $i-1$ is associated with a parent, at level $i$, which lies lies within distance $2^i$.

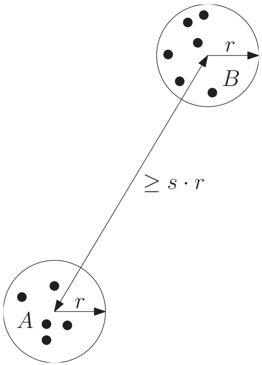


# Well-Separated Pair Decompositions (WSPD)

## Well-Separated Pair Decomposition

- $n$ points determine $O(n^2)$ pairs
- $A$ and $B$ are $s$-well separated if they can be enclosed in balls of radius $r$ that are separated by at least $s \cdot r$
- A WSPD of a point set $P$ is a collection of well-separated pairs $(A_i, B_i)$ covering all pairs of the set
- An $n$-element point set in dimension $d$ has a WSPD of size $O(s^d n) = O(n)$ [CaK95]



# Computational Issues

## Main Computational Issues

- Spread: $-\sum_{i \neq j} \log\left(1 + e^{\eta_{i,j}}\right)$
- Stretch: $\sum_{i \neq j} \eta_{i,j} y_{i,j}$
- Clustered Motion: Moving blocks of points efficiently
- Dynamics: Updating the data structures

# Computational Issues

## Main Computational Issues

- Spread: $-\sum_{i \neq j} \log\left(1 + e^{\eta_{i,j}}\right) \;\Rightarrow\;$ Locally sensitive sampling
- Stretch: $\sum_{i \neq j} \eta_{i,j} y_{i,j}$
- Clustered Motion: Moving blocks of points efficiently
- Dynamics: Updating the data structures

# Computational Issues

## Main Computational Issues

- Spread: $-\sum_{i \neq j} \log\left(1 + e^{\eta_{i,j}}\right) \Rightarrow$ Locally sensitive sampling
- Stretch: $\sum_{i \neq j} \eta_{i,j} y_{i,j} \Rightarrow$ Power-series expansion
- Clustered Motion: Moving blocks of points efficiently
- Dynamics: Updating the data structures

# Computational Issues

## Main Computational Issues

- Spread: $-\sum_{i \neq j} \log\left(1 + e^{\eta_{i,j}}\right) \Rightarrow$ Locally sensitive sampling
- Stretch: $\sum_{i \neq j} \eta_{i,j} y_{i,j} \Rightarrow$ Power-series expansion
- Clustered Motion: Moving blocks of points efficiently $\Rightarrow$ WSPDs
- Dynamics: Updating the data structures

# Computational Issues – Spread

Spread Term:

$$-\sum_{i \neq j} \log\left(1 + e^{\alpha - \|z_i - z_j\|}\right)$$

- Independent of edges
- Dominated by nearby objects
  (Tends quickly to zero as $\|z_i - z_j\|$ increases)
- Proposed Approach: Locally sensitive sampling:
  - Compute a WSPD with a low separation factor
  - This provides a crude estimate of the distance distribution
  - Sample pairs at random, favoring pairs that are close

# Computational Issues – Stretch

Stretch Term: $$\sum_{i \neq j} \eta_{i,j} y_{i,j} = \sum_{(i,j) \in E} (\alpha - \|z_i - z_j\|) = \alpha|E| - \sum_{(i,j) \in E} \|z_i - z_j\|.$$

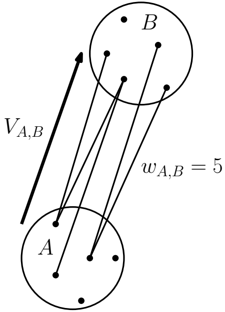
- Computable in time proportional to the number of edges
- Sparse Graphs: ($|E| = O(n)$) Compute by brute force
- Dense Graphs: ($|E| \gg O(n)$)
  - Euclidean Distance: Approximate through a combination of power-series expansion and WSPDs (as in FMM)
  - Squared Euclidean Distance: Efficient block motion

# Computational Issues – Stretch with Squared Distances

Stretch with Squared Distances: $\alpha|E| - \sum_{(i,j)\in E} \|z_i - z_j\|^2$

## Preprocessing

- Build a WSPD $\Phi = \{(A_1, B_1), (A_2, B_2), \ldots\}$.
- For each pair $(A, B) \in \Phi$, let $E_{A,B} = |E \cap (A \times B)|$. Maintain:
  - Weight: $w_{A,B} = |E_{A,B}|$
  - Centroid Displacement Vector: $V_{A,B} = \frac{1}{w_{A,B}} \sum_{(a,b)\in E_{A,B}} (b - a)$
  - Base Stretch: $\Delta_{A,B} = \frac{1}{w_{A,B}} \sum_{(a,b)\in E_{A,B}} \|b - a\|^2$



# Computational Issues – Stretch with Squared Distances

Stretch with Squared Distances: $\alpha|E| - \sum_{(i,j)\in E} \|z_i - z_j\|^2$

## Block-Motion Update

If $B$ is translated by $t$ relative to $A$, then can update $\Delta_{A,B}$ in $O(1)$ time.

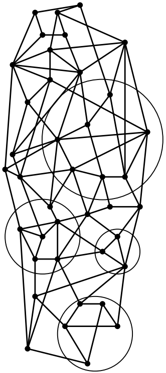
$$\begin{aligned}
\Delta_{A,B+t} &= \frac{1}{w_{A,B}} \sum_{(a,b)} \|(b+t) - a\|^2 = \frac{1}{w_{A,B}} \sum_{(a,b)} \|t + (b-a)\|^2 \\
&= \frac{1}{w_{A,B}} \sum_{(a,b)} (t \cdot t) + 2(t \cdot (b-a)) + (b-a) \cdot (b-a) \\
&= \frac{1}{w_{A,B}} \left( w_{A,B}(t \cdot t) + 2(t \cdot \sum_{(a,b)} (b-a)) + \sum_{(a,b)} \|b-a\|^2 \right) \\
&= (t \cdot t) + 2(t \cdot V_{A,B}) + \Delta_{A,B}.
\end{aligned}$$

# Computational Issues – Hierarchical Block Motion

## Hierarchical Block-Motion

If squared distances are used, we can move $k$ blocks of points in $O(k)$ time.

- Use the net tree to define blocks at various resolutions.
- WSPD and associated values, $w_{A,B}$, $V_{A,B}$, $\Delta_{A,B}$ are maintained in the net tree.
- Updates to block membership can be performed efficiently in $O(\log n)$ time.
- Standard Euclidean Distances: Can approximate using power series.

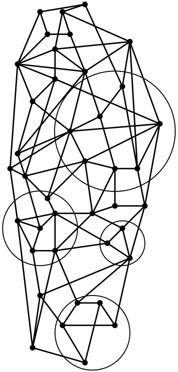


# Computational Issues – Hierarchical Block Motion

## Hierarchical Block-Motion

If squared distances are used, we can move $k$ blocks of points in $O(k)$ time.

- Use the net tree to define blocks at various resolutions.
- WSPD and associated values, $w_{A,B}$, $V_{A,B}$, $\Delta_{A,B}$ are maintained in the net tree.
- Updates to block membership can be performed efficiently in $O(\log n)$ time.
- Standard Euclidean Distances: Can approximate using power series.



# Future Work

- Continue to refine computational methods
- Prototype algorithms and data structures
- Empirical analysis of accuracy and efficiency

Thank you!

# Bibliography


- [CK95] P. B. Callahan and S. R. Kosaraju. A decomposition of multidimensional point sets with applications to $k$-nearest-neighbors and $n$-body potential fields. *J. Assoc. Comput. Mach.*, 42:67–90, 1995.
- [HRH02] P. D. Hoff, A. E. Raftery, and M. S Handcock. Latent space approaches to social network analysis. *J. American Statistical Assoc.*, 97:1090–1098, 2002.
- [HRT07] M. S. Handcock and A. E. Raftery and J. M. Tantrum. Model-based clustering for social networks. *J. R. Statist. Soc. A*, 170, Part 2, 301–354, 2007.