On Clustering Using Random Walks

David Harel and Yehuda Koren

presented at Foundations of Software Technology and Theoretical Computer Science 2001

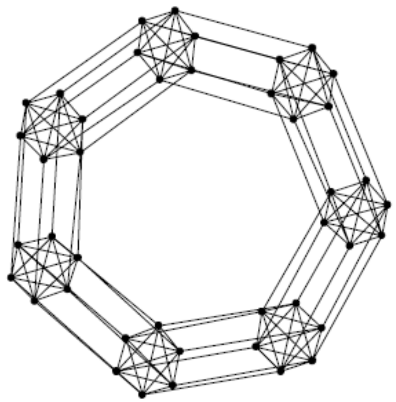
Talk by Kota Hara

OVERVIEW

- Important properties
- Method
  - Flow of the method
  - Creating structured graph
  - Random walk
  - Separating operator
  - Integration with agglomerative clustering
- Results
- Summary

PROPERTIES OF METHOD

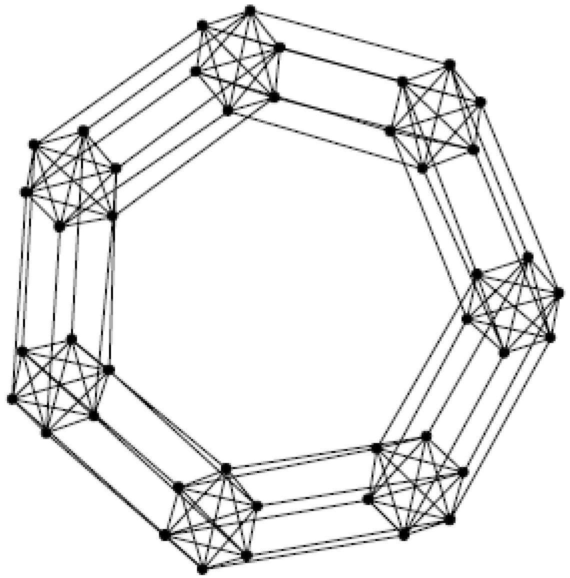
- Exploit structure of graph
- Based on deterministic analysis of random walks on weighted graphs (no random simulation)
- Doesn't require a prespecified number of cluster
- No eigenvalue problem involved
- Work well on clustering of 2d spatial points




DS4: 8000 points
DS5: 8000 points
DS6: 10000 points

How it works? (Simple example)

- Want to divide points into 7 clusters
- Create structured graph
- 'Separating operator' decreases weights of 'external edge' while increases those of 'internal edge'
- Remove edges with small weights
- Or apply agglomerative clustering



Creating structured graph $G(V, E, w)$

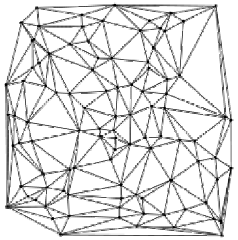
Given data points….

Weight

- $exp(-\frac{d(a,b)^2}{C})$ when clustering spatial points
- Uniform in some examples

Edge

- k-mutual neighborhood
- Remove edges whose weights are below threshold
- Delaunay triangulation



Random walk on graph

- 'Random walker' walks on graph according to probability of transition (Markov process)

- Probability of transition from node i to j

$$p_{ij} = \frac{w(i,j)}{d_i} \quad \text{where} \quad d_i = \sum_{\langle i,k \rangle} w(i,k)$$

- Transition matrix

$$M_{ij} = \begin{cases} p_{ij} & \langle i,j \rangle \in E \\ 0 & otherwise \end{cases}$$

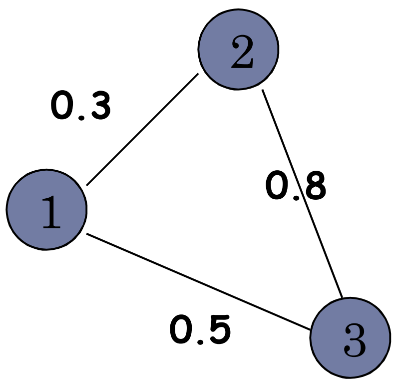
TRANSITION MATRIX

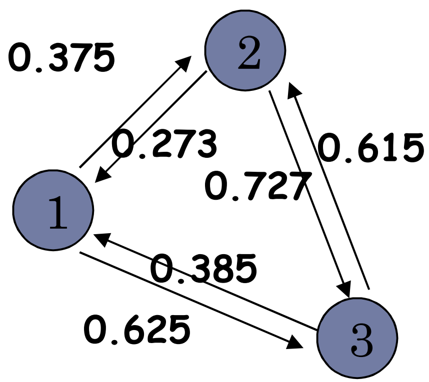
0	0.3	0.5
0.3	0	0.8
0.5	0.8	0

Similarity matrix A

0	0.375	0.625
0.273	0	0.727
0.385	0.615	0

Transition matrix *M*





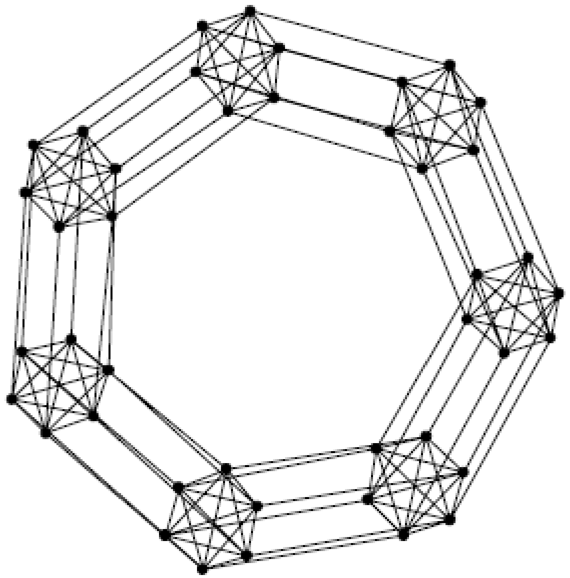
Transition probability

- Vector whose j-th component is probability that a random walk originating at i will visit node j in its k-th step $P^k_{visit}(i) \in \mathbb{R}^n$
- $P^k_{visit}(i)$ is the i-th row in M^k

Separating operator

- Based on deterministic analysis of random walks

- Applied to the graph several times to increase weights of intra-cluster edges and decrease inter-cluster edges



SEPARATING OPERATOR 1 (NS)

- Vector whose j-th component is probability that a random walk originating at i will visit node j within k steps (k=3)

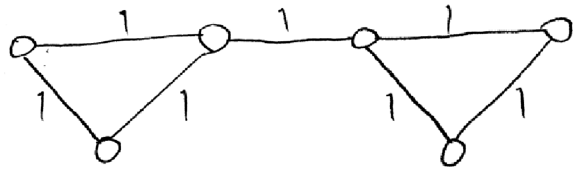
$$P_{visit}^{\leq k}(i) = \sum_{s=1}^{k} P_{visit}^{s}(i)$$

- Use similarity between $P_{visit}^{\leq k}(v)$ and $P_{visit}^{\leq k}(u)$ as a similarity between v and u

- Replace $w(u,v) \quad \forall \langle u,v \rangle \in E$ by

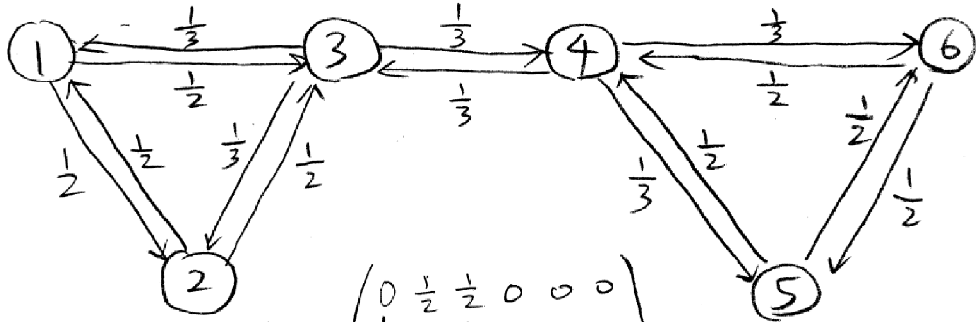
$$\exp(2k - \|P_{visit}^{\leq k}(v) - P_{visit}^{\leq k}(u)\|_{L_1}) - 1$$

weights w



$\Rightarrow$

transition probability



$$M = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \rightarrow 1$$

$$\left.\begin{array}{l} P^{1}_{visit}(4) = \left(0 \;\; 0 \;\; \frac{1}{3} \;\; 0 \;\; \frac{1}{3} \;\; \frac{1}{3}\right) \\ P^{2}_{visit}(4) = \left(\frac{1}{9} \;\; \frac{1}{9} \;\; 0 \;\; \frac{4}{9} \;\; \frac{1}{6} \;\; \frac{1}{6}\right) \end{array}\right\} \; P^{k\le 2}_{visit}(4) = \left(\frac{1}{9} \;\; \frac{1}{9} \;\; \frac{1}{3} \;\; \frac{4}{9} \;\; \frac{1}{2} \;\; \frac{1}{2}\right)$$

$$\left.\begin{array}{l} P^{1}_{visit}(5) = \left(0 \;\; 0 \;\; 0 \;\; \frac{1}{2} \;\; 0 \;\; \frac{1}{2}\right) \\ P^{2}_{visit}(5) = \left(0 \;\; 0 \;\; \frac{1}{6} \;\; \frac{1}{4} \;\; \frac{5}{12} \;\; \frac{1}{6}\right) \end{array}\right\} \; P^{k\le 2}_{visit}(5) = \left(0 \;\; 0 \;\; \frac{1}{6} \;\; \frac{3}{4} \;\; \frac{5}{12} \;\; \frac{2}{3}\right)$$

$$\left.\begin{array}{l} P^{1}_{visit}(3) = \left(\frac{1}{3} \;\; \frac{1}{3} \;\; 0 \;\; \frac{1}{3} \;\; 0 \;\; 0\right) \\ P^{2}_{visit}(3) = \left(\frac{1}{6} \;\; \frac{1}{6} \;\; \frac{4}{9} \;\; 0 \;\; \frac{1}{9} \;\; \frac{1}{9}\right) \end{array}\right\} \; P^{k\le 2}_{visit}(3) = \left(\frac{1}{2} \;\; \frac{1}{2} \;\; \frac{4}{9} \;\; \frac{1}{3} \;\; \frac{1}{9} \;\; \frac{1}{9}\right)$$

$$\left\| P^{k\le 2}_{visit}(4) - P^{k\le 2}_{visit}(5) \right\|_{L1} = \frac{17}{18} \quad \Longrightarrow \quad w(4,5) = 20{,}233$$

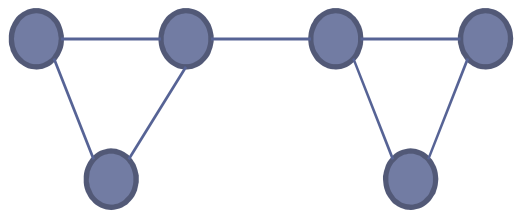
$$\left\| P^{k\le 2}_{visit}(4) - P^{k\le 2}_{visit}(3) \right\|_{L1} = \frac{32}{18} \quad \Longrightarrow \quad w(3,4) = 0{,}2278$$

SEPARATING OPERATOR 2 (CE)

- Probability that random walker starting from v visits u before returning to u (escape probability)

$$P_{escape}(v, u)$$

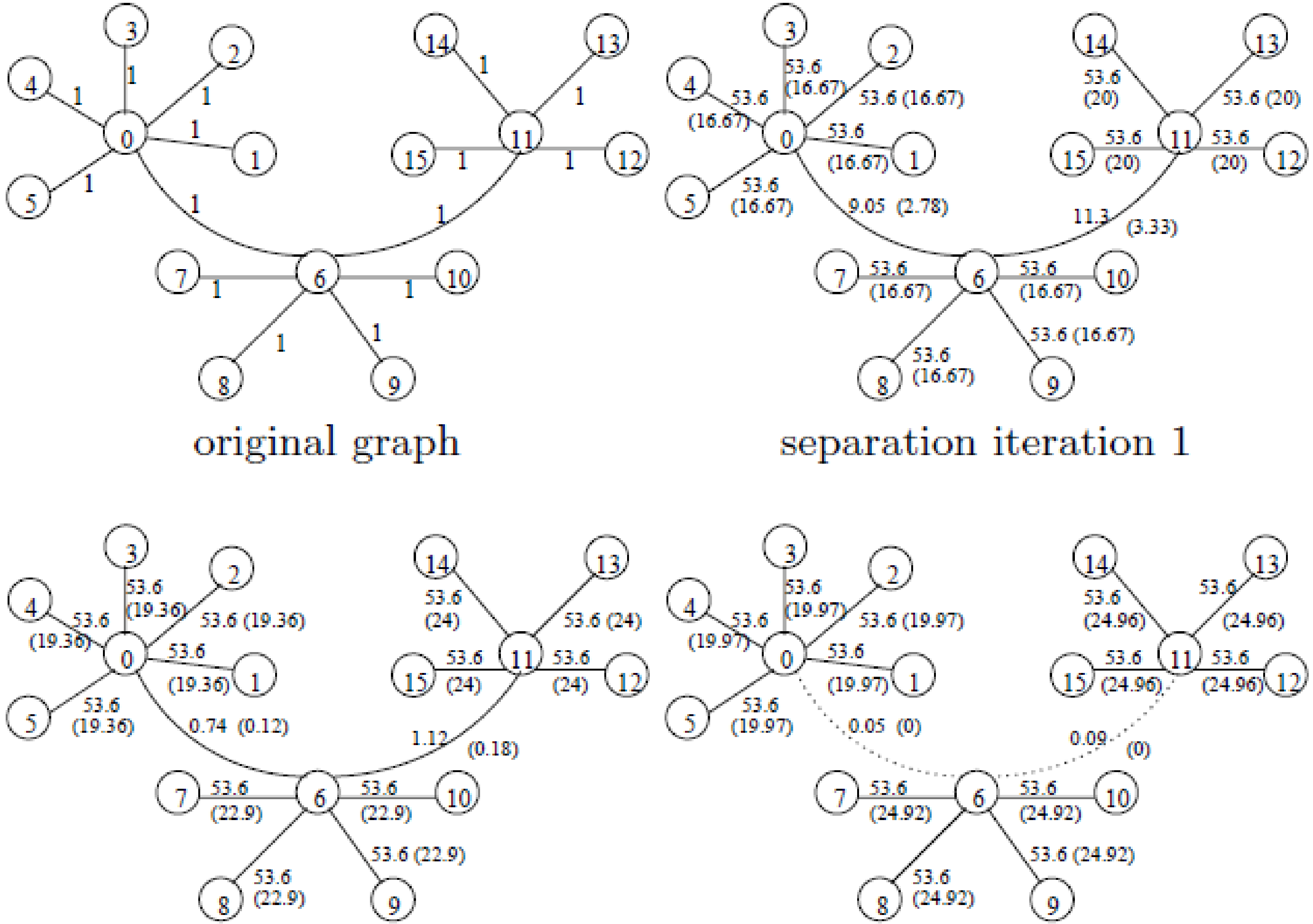
- If v and u are in different cluster, $P_{escape}(v, u)$ will be low



- Replace $w(u, v) \quad \forall \langle u, v \rangle \in E$ by

$$P_{escape}(v, u) \cdot P_{escape}(u, v)$$

EXAMPLE



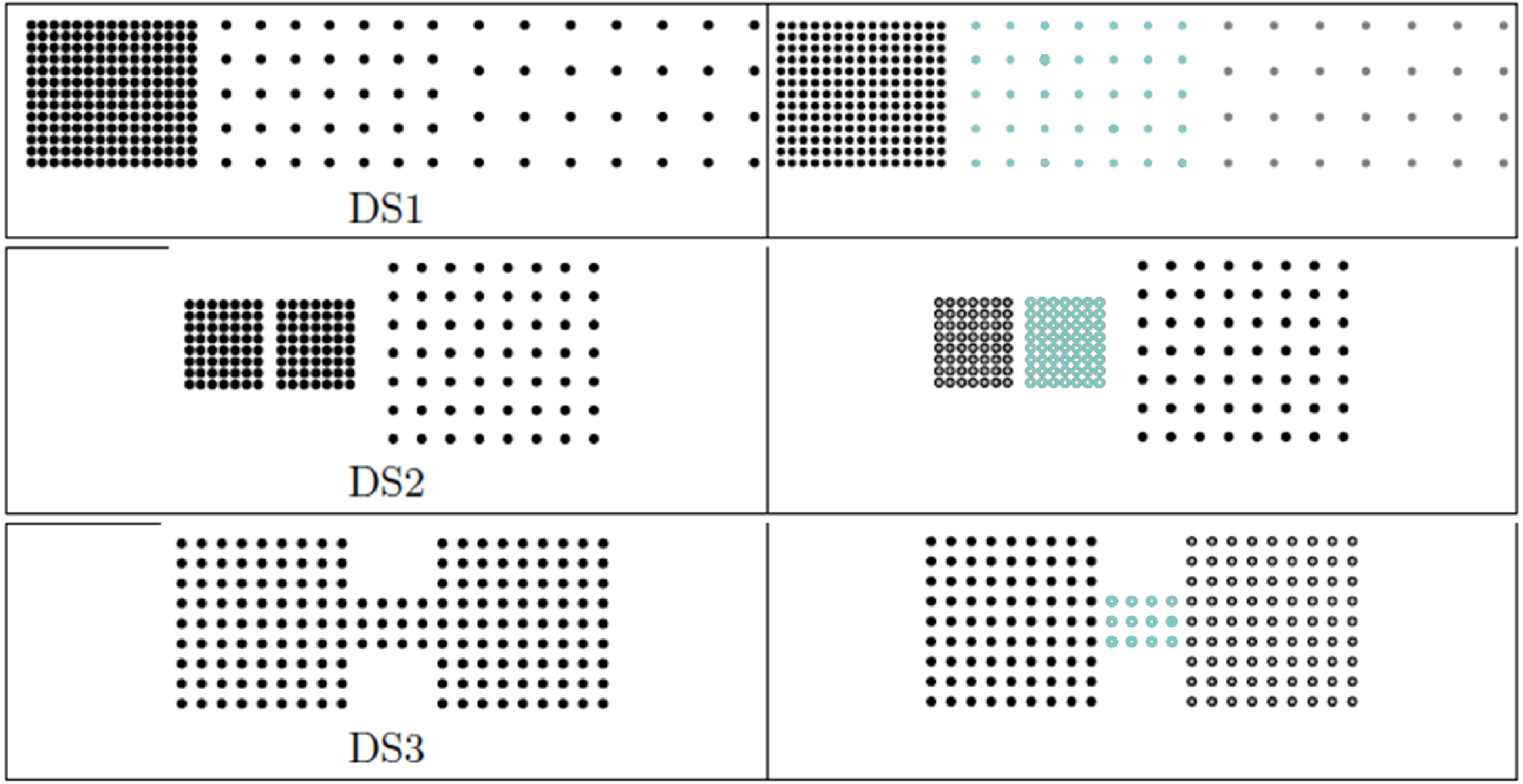
original graph

separation iteration 1

separation iteration 2

separation iteration 3

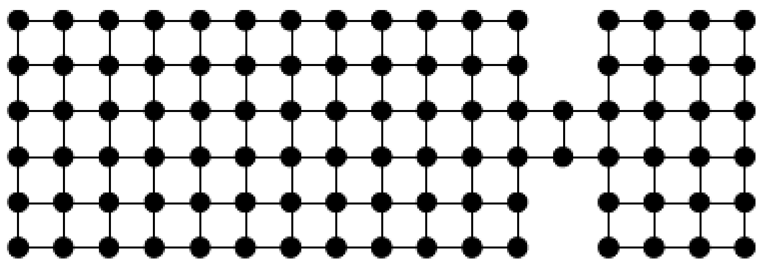
Clustering spatial point-sets



Integration with agglomerative clustering

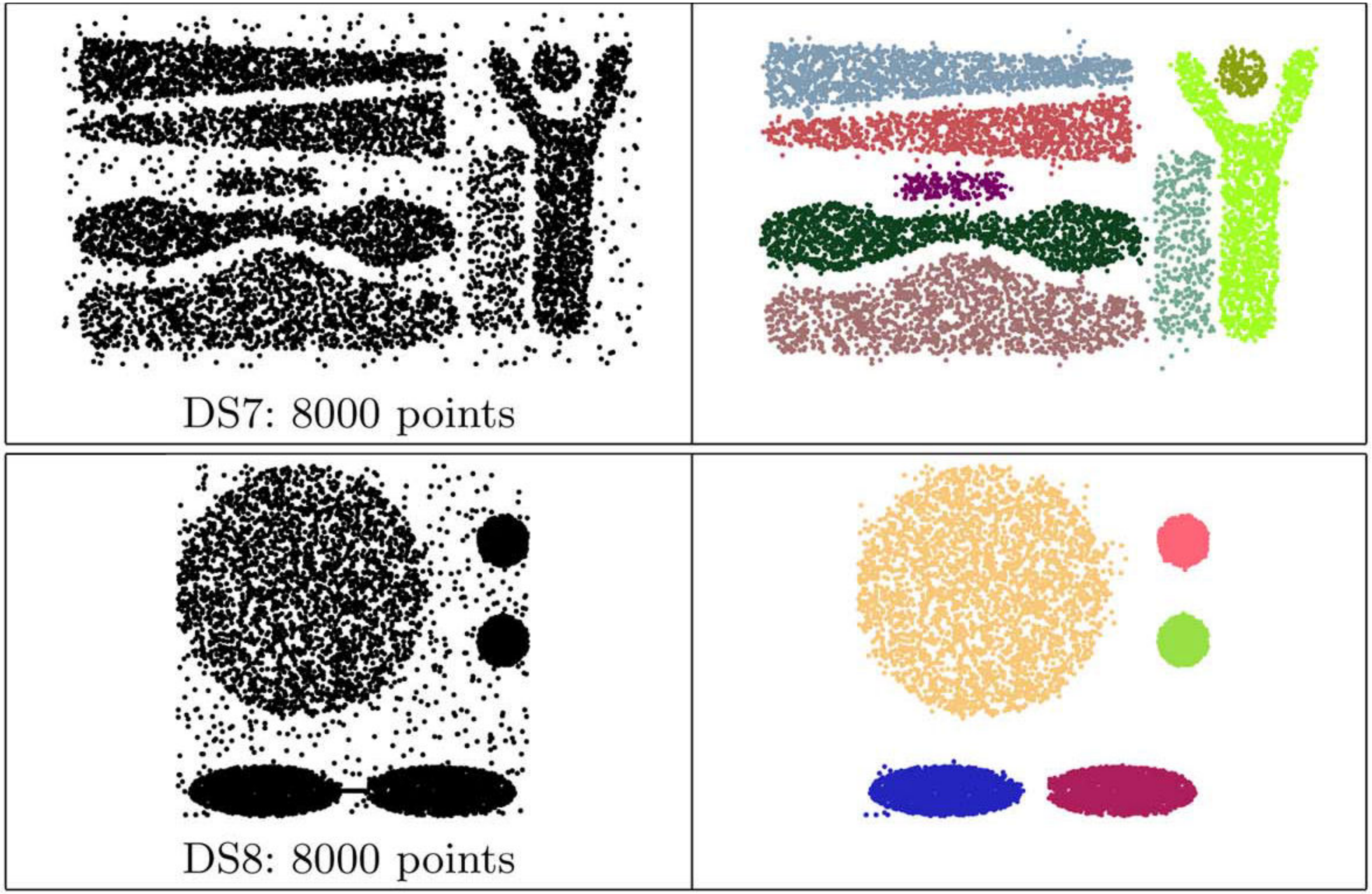
- Use separation operation as a preprocessing stage before agglomerative clustering

- Prevent bad local merging that works against the graph structure



- (Variant) Repeatedly apply separation operation during agglomerative clustering

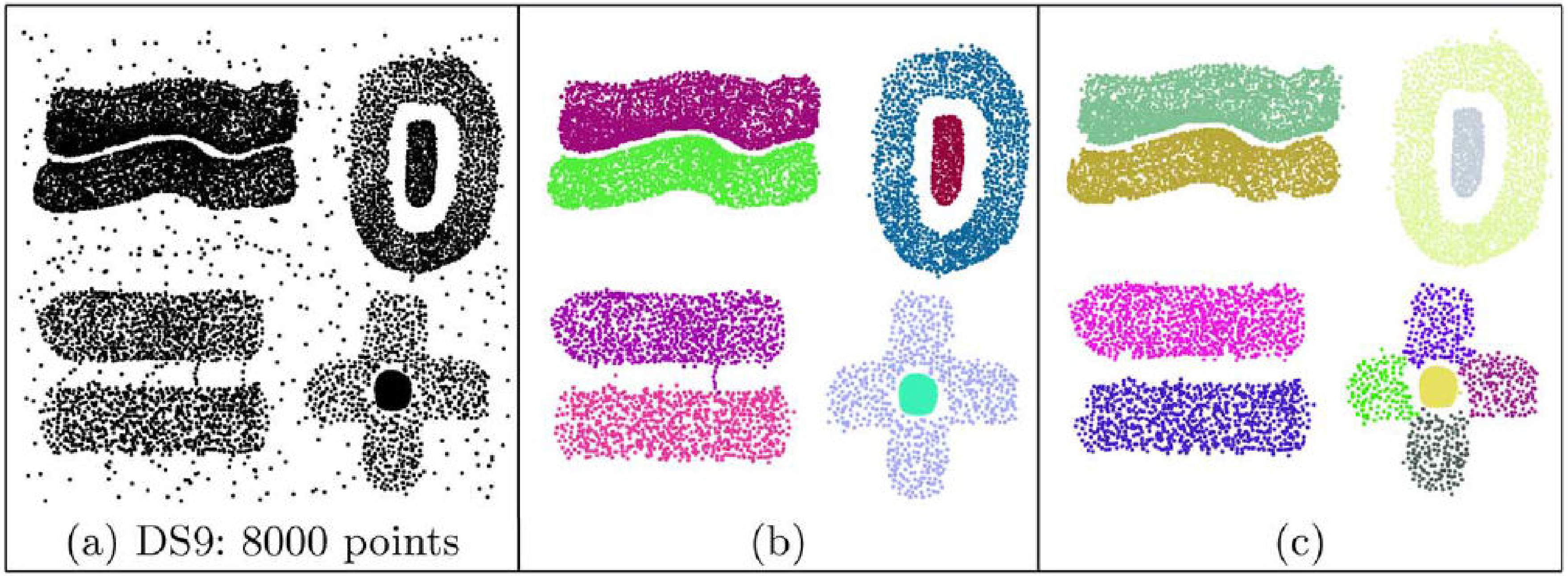
Results



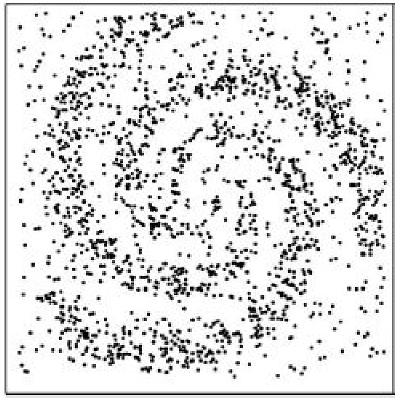
DS7: 8000 points

DS8: 8000 points

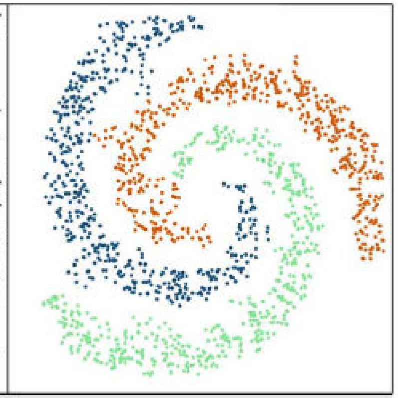
RESULTS (TWO LEVELS IN HIERARCHY)



COMPARATIVE RESULTS

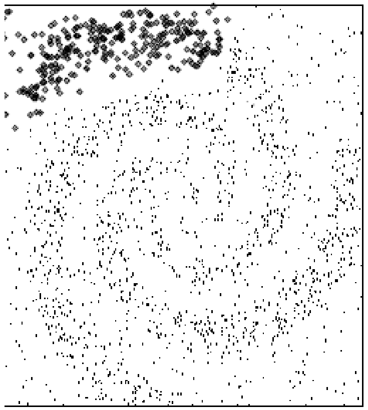


Input

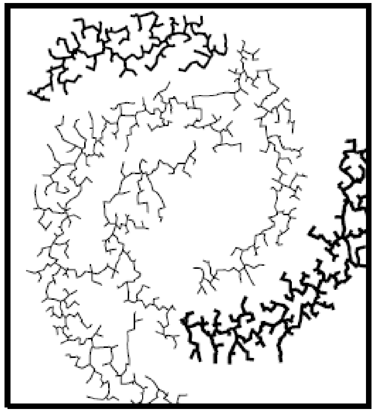


Their method

separating
operator
+
agglomerative
clustering



Normalized
cut



Agglomerative
clustering

SUMMARY

- Deterministic analysis of random walks on weighted graph (does not use random algorithms to simulate random walks)
- Exploit structure of graph in addition to weights
- Separating operators 'sharpen' the distinction between weights of inter-cluster edge and intra-cluster edge
- More powerful when incorporated with agglomerative clustering

Reference

- On Clustering Using Random Walks, David Harel and Yehuda Koren, FSTTCS 2011
- Stochastic Image Segmentation by Typical Cuts, Fort Collins, Colorado, CVPR 1999
- CHAMELEON: A Hierarchical Clustering Algorithm Using Dynamic Modeling, George Karypis , Eui-Hong (Sam) Han , Vipin Kumar, IEEE Computer 32(1999), 68-75
- http://stat-www.berkeley.edu/users/aldous/RWG/book.html

Question?

Computation time

Table 1. Running time (in seconds; non-optimized) of the various components of the clustering algorithm

Data Set	Size	Graph construction	Separation	Agglomeration	Overall	Ratio $\frac{Points}{Sec}$
DS4	8000	0.4	0.88	0.19	1.47	5434
DS5	8000	0.41	0.83	0.19	1.43	5587
DS6	10000	0.5	1.12	0.26	1.88	5311
DS7	8000	0.4	0.89	0.2	1.49	5358
DS8	8000	0.39	0.93	0.2	1.52	5256
DS9	8000	0.33	0.66	0.21	1.2	6656
DS10	3374	0.14	0.26	0.07	0.47	7178

700MHz Pentium III PC

Basic notation

- Weighted graph $G(V, E, w)$
- Set of nodes connected to S by a path with at most k edges $V^k(S)$
- Maxima number of edges incident to some single node of G $deg(G)$
- Subgraph of G induced by S

$$G(S)$$