Script generated by TTT

Title: groh: profile1 (10.06.2015)

Date: Wed Jun 10 08:17:02 CEST 2015

Duration: 79:14 min

Pages: 64

Social Gaming / Social Computing SS 2015

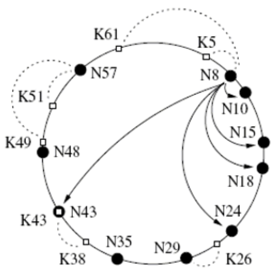
PD Dr. Georg Groh

Applied Informatics / Cooperative Systems AICOS

Technische Universität München

Homophily and Distance

- Example for this principle: Early P2P file sharing protocols based on Distributed Hash Tables: Example: Chord:
  - data (e.g. filenames) („keys") and host-IDs (e.g. IP-addresses) („nodes") hashed into the same m-dim key-space,
  - Key *k* is assigned to node *successor(k)*,
  - each node n maintains finger table (length up to m): m-th entry: IP-address of node *successor((n + 2^{i-1}) mod 2^m)* → routing in log(N) steps possible



Finger Table of N8

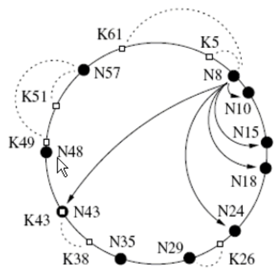
Idx	Target ID	Successor
0	N8 + 1	N10
1	N8 + 2	N10
2	N8 + 4	N15
3	N8 + 8	N18
4	N8 + 16	N24
5	N8 + 32	N43

Fig. 8.1: A 6-bit Chord identifier space. Dotted lines indicate which nodes host which keys. Black lines represent the fingers of node *N*8.

[8]

Homophily and Distance

- Example for this principle: Early P2P file sharing protocols based on Distributed Hash Tables: Example: Chord:
  - data (e.g. filenames) („keys") and host-IDs (e.g. IP-addresses) („nodes") hashed into the same m-dim key-space,
  - Key *k* is assigned to node *successor(k)*,
  - each node n maintains finger table (length up to m): m-th entry: IP-address of node *successor((n + 2^{i-1}) mod 2^m)* → routing in log(N) steps possible



Finger Table of N8

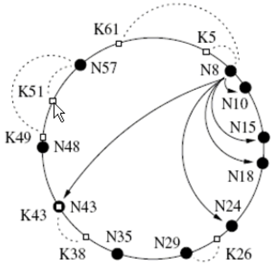
Idx	Target ID	Successor
0	N8 + 1	N10
1	N8 + 2	N10
2	N8 + 4	N15
3	N8 + 8	N18
4	N8 + 16	N24
5	N8 + 32	N43

Fig. 8.1: A 6-bit Chord identifier space. Dotted lines indicate which nodes host which keys. Black lines represent the fingers of node *N*8.

[8]

Homophily and Distance

- Example for this principle: Early P2P file sharing protocols based on Distributed Hash Tables: Example: Chord:
  - data (e.g. filenames) („keys“) and host-IDs (e.g. IP-addresses) („nodes“) hashed into the same m-dim key-space,
  - Key *k* is assigned to node *successor(k)*,
  - each node n maintains finger table (length up to m): m-th entry: IP-address of node *successor*$((n + 2^{i-1}) \bmod 2^m)$ → routing in log(N) steps possible



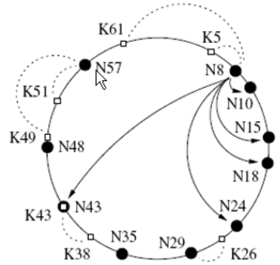
Finger Table of N8		
Idx	Target ID	Successor
0	N8 + 1	N10
1	N8 + 2	N10
2	N8 + 4	N15
3	N8 + 8	N18
4	N8 + 16	N24
5	N8 + 32	N43

Fig. 8.1: A 6-bit Chord identifier space. Dotted lines indicate which nodes host which keys. Black lines represent the fingers of node *N*8.

[8]

Homophily and Distance

- Example for this principle: Early P2P file sharing protocols based on Distributed Hash Tables: Example: Chord:
  - data (e.g. filenames) („keys“) and host-IDs (e.g. IP-addresses) („nodes“) hashed into the same m-dim key-space,
  - Key *k* is assigned to node *successor(k)*,
  - each node n maintains finger table (length up to m): m-th entry: IP-address of node *successor*$((n + 2^{i-1}) \bmod 2^m)$ → routing in log(N) steps possible



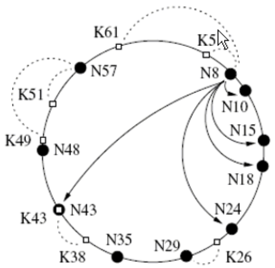
Finger Table of N8		
Idx	Target ID	Successor
0	N8 + 1	N10
1	N8 + 2	N10
2	N8 + 4	N15
3	N8 + 8	N18
4	N8 + 16	N24
5	N8 + 32	N43

Fig. 8.1: A 6-bit Chord identifier space. Dotted lines indicate which nodes host which keys. Black lines represent the fingers of node *N*8.

[8]

Homophily and Distance

- Example for this principle: Early P2P file sharing protocols based on Distributed Hash Tables: Example: Chord:
  - data (e.g. filenames) („keys“) and host-IDs (e.g. IP-addresses) („nodes“) hashed into the same m-dim key-space,
  - Key *k* is assigned to node *successor(k)*,
  - each node n maintains finger table (length up to m): m-th entry: IP-address of node *successor*$((n + 2^{i-1}) \bmod 2^m)$ → routing in log(N) steps possible



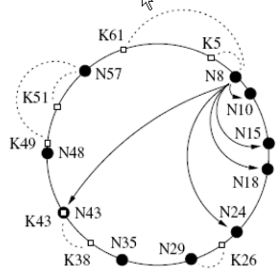
Finger Table of N8		
Idx	Target ID	Successor
0	N8 + 1	N10
1	N8 + 2	N10
2	N8 + 4	N15
3	N8 + 8	N18
4	N8 + 16	N24
5	N8 + 32	N43

Fig. 8.1: A 6-bit Chord identifier space. Dotted lines indicate which nodes host which keys. Black lines represent the fingers of node *N*8.

[8]

Homophily and Distance

- Example for this principle: Early P2P file sharing protocols based on Distributed Hash Tables: Example: Chord:
  - data (e.g. filenames) („keys“) and host-IDs (e.g. IP-addresses) („nodes“) hashed into the same m-dim key-space,
  - Key *k* is assigned to node *successor(k)*,
  - each node n maintains finger table (length up to m): m-th entry: IP-address of node *successor*$((n + 2^{i-1}) \bmod 2^m)$ → routing in log(N) steps possible



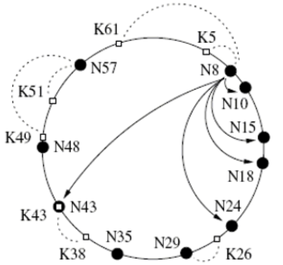
Finger Table of N8		
Idx	Target ID	Successor
0	N8 + 1	N10
1	N8 + 2	N10
2	N8 + 4	N15
3	N8 + 8	N18
4	N8 + 16	N24
5	N8 + 32	N43

Fig. 8.1: A 6-bit Chord identifier space. Dotted lines indicate which nodes host which keys. Black lines represent the fingers of node *N*8.

[8]

Homophily and Distance

- Example for this principle: Early P2P file sharing protocols based on Distributed Hash Tables: Example: Chord:
  - data (e.g. filenames) („keys") and host-IDs (e.g. IP-addresses) („nodes") hashed into the same m-dim key-space,
  - Key *k* is assigned to node *successor(k)*,
  - each node n maintains finger table (length up to m): m-th entry: IP-address of node *successor*$((n + 2^{i-1}) \bmod 2^m)$ → routing in log(N) steps possible



Finger Table of N8

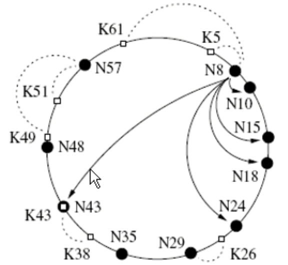
Idx	Target ID	Successor
0	N8 + 1	N10
1	N8 + 2	N10
2	N8 + 4	N15
3	N8 + 8	N18
4	N8 + 16	N24
5	N8 + 32	N43

Fig. 8.1: A 6-bit Chord identifier space. Dotted lines indicate which nodes host which keys. Black lines represent the fingers of node $N8$.

[8]

Homophily and Distance

- Example for this principle: Early P2P file sharing protocols based on Distributed Hash Tables: Example: Chord:
  - data (e.g. filenames) („keys") and host-IDs (e.g. IP-addresses) („nodes") hashed into the same m-dim key-space,
  - Key *k* is assigned to node *successor(k)*,
  - each node n maintains finger table (length up to m): m-th entry: IP-address of node *successor*$((n + 2^{i-1}) \bmod 2^m)$ → routing in log(N) steps possible



Finger Table of N8

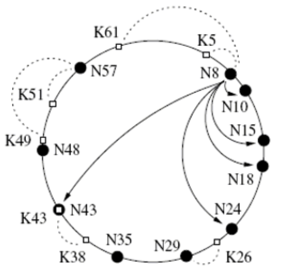
Idx	Target ID	Successor
0	N8 + 1	N10
1	N8 + 2	N10
2	N8 + 4	N15
3	N8 + 8	N18
4	N8 + 16	N24
5	N8 + 32	N43

Fig. 8.1: A 6-bit Chord identifier space. Dotted lines indicate which nodes host which keys. Black lines represent the fingers of node $N8$.

[8]

Homophily and Distance

- Example for this principle: Early P2P file sharing protocols based on Distributed Hash Tables: Example: Chord:
  - data (e.g. filenames) („keys") and host-IDs (e.g. IP-addresses) („nodes") hashed into the same m-dim key-space,
  - Key *k* is assigned to node *successor(k)*,
  - each node n maintains finger table (length up to m): m-th entry: IP-address of node *successor*$((n + 2^{i-1}) \bmod 2^m)$ → routing in log(N) steps possible



Finger Table of N8

Idx	Target ID	Successor
0	N8 + 1	N10
1	N8 + 2	N10
2	N8 + 4	N15
3	N8 + 8	N18
4	N8 + 16	N24
5	N8 + 32	N43

Fig. 8.1: A 6-bit Chord identifier space. Dotted lines indicate which nodes host which keys. Black lines represent the fingers of node $N8$.

[8]

Kleinberg Model

- Put nodes on a n x n grid. Distance: Manhattan:

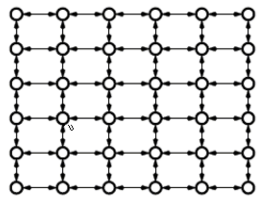
$$r(i,j) = |x_i - x_j| + |y_i - y_j|$$

- Each node i: Connected to all nodes with $r(i,j) \leq q1$ (regular local contacts)
- Each node i: Additional q2 other „long range" edges: Probability of edge to node j:

$$P(j) \sim r(i,j)^{-\alpha}$$

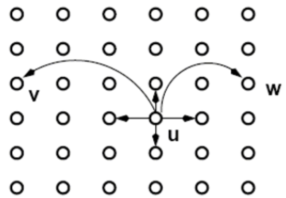
examples:

A)



q1 = 1, q2 = 0

B)



edges for one node u: (q1 = 1, q2 = 2)

[3]

Kleinberg Model

- Put nodes on a n x n grid. Distance: Manhattan:

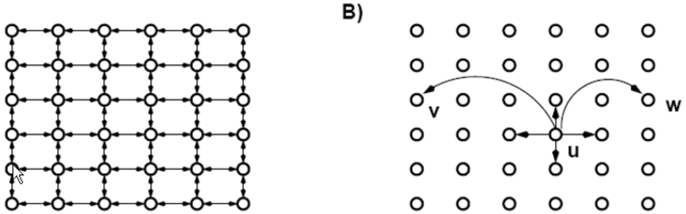
$$r(i,j) = |x_i - x_j| + |y_i - y_j|$$

- Each node i: Connected to all nodes with $r(i,j) \leq q1$ (regular local contacts)
- Each node i: Additional q2 other „long range" edges:
  Probability of edge to node j:

$$P(j) \sim r(i,j)^{-\alpha}$$

examples:

A) B)



q1 = 1, q2 = 0 edges for one node u: (q1 = 1, q2 = 2)

[3]

Kleinberg Model

- Put nodes on a n x n grid. Distance: Manhattan:

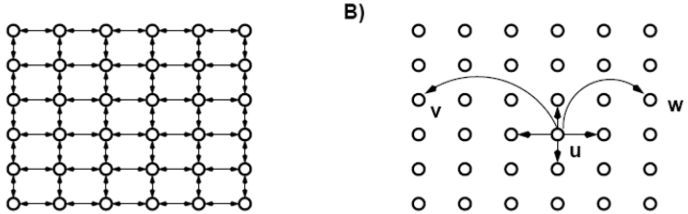
$$r(i,j) = |x_i - x_j| + |y_i - y_j|$$

- Each node i: Connected to all nodes with $r(i,j) \leq q1$ (regular local contacts)
- Each node i: Additional q2 other „long range" edges:
  Probability of edge to node j:

$$P(j) \sim r(i,j)^{-\alpha}$$

examples:

A) B)



q1 = 1, q2 = 0 edges for one node u: (q1 = 1, q2 = 2)

[3]

Kleinberg Model

- Put nodes on a n x n grid. Distance: Manhattan:

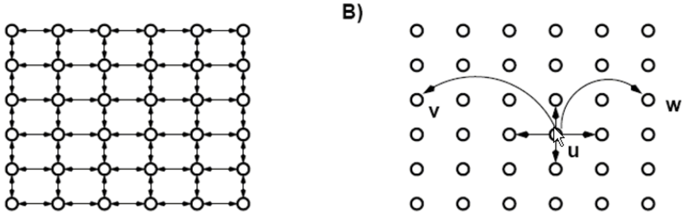
$$r(i,j) = |x_i - x_j| + |y_i - y_j|$$

- Each node i: Connected to all nodes with $r(i,j) \leq q1$ (regular local contacts)
- Each node i: Additional q2 other „long range" edges:
  Probability of edge to node j:

$$P(j) \sim r(i,j)^{-\alpha}$$

examples:

A) B)



q1 = 1, q2 = 0 edges for one node u: (q1 = 1, q2 = 2)

[3]

Kleinberg Model

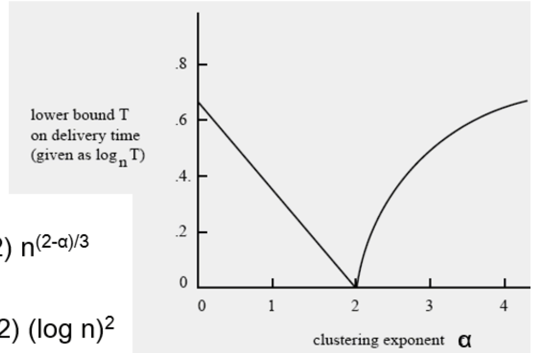
- local (decentralized) knowledge:
  - Each node only knows only:
    - Its adjacent nodes
    - The grid's principle structure
    - Position of target node on the grid
    - Positions and long-range contacts of nodes on the message path so far
- (Search-) algorithm with only local knowledge: „decentralized"

Kleinberg Model

- local (decentralized) knowledge:
  - Each node only knows only:
    - Its adjacent nodes
    - The grid‘s principle structure
    - Position of target node on the grid
    - Positions and long-range contacts of nodes on the message path so far
- (Search-) algorithm with only local knowledge: „decentralized“

Kleinberg Model

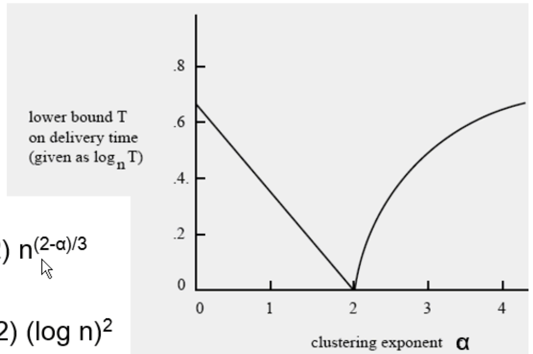
- Now: Send message with local (decentralized) knowledge only
- Given: Decentralized greedy message delivery algorithm: measure number of expected delivery steps s:
  - $0 \le \alpha < 2$: s at least $\sim c1(\alpha,q1,q2)\ n^{(2-\alpha)/3}$
  - $\alpha = 2$: s at most $\sim c2(\alpha,q1,q2)\ (\log n)^2$
  - $\alpha > 2$: s at least $\sim c3(\alpha,q1,q2)\ n^{(\alpha-2)/(\alpha-1)}$



(closely after [3])

Kleinberg Model

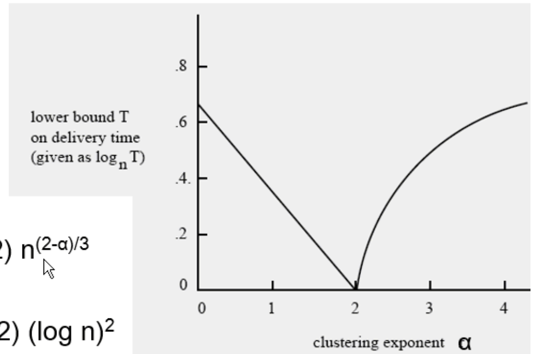
- Now: Send message with local (decentralized) knowledge only
- Given: Decentralized greedy message delivery algorithm: measure number of expected delivery steps s:
  - $0 \le \alpha < 2$: s at least $\sim c1(\alpha,q1,q2)\ n^{(2-\alpha)/3}$
  - $\alpha = 2$: s at most $\sim c2(\alpha,q1,q2)\ (\log n)^2$
  - $\alpha > 2$: s at least $\sim c3(\alpha,q1,q2)\ n^{(\alpha-2)/(\alpha-1)}$



(closely after [3])

Kleinberg Model

- Now: Send message with local (decentralized) knowledge only
- Given: Decentralized greedy message delivery algorithm: measure number of expected delivery steps s:
  - $0 \le \alpha < 2$: s at least $\sim c1(\alpha,q1,q2)\ n^{(2-\alpha)/3}$
  - $\alpha = 2$: s at most $\sim c2(\alpha,q1,q2)\ (\log n)^2$
  - $\alpha > 2$: s at least $\sim c3(\alpha,q1,q2)\ n^{(\alpha-2)/(\alpha-1)}$



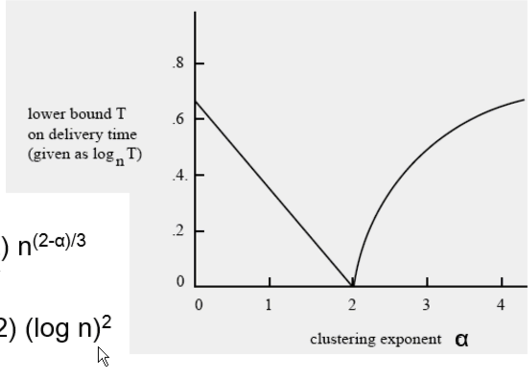
(closely after [3])

Kleinberg Model

- Now: Send message with local (decentralized) knowledge only

- Given: Decentralized greedy message delivery algorithm: measure number of expected delivery steps s:

  - $0 \le \alpha < 2$: s at least $\sim c1(\alpha,q1,q2)\ n^{(2-\alpha)/3}$
  - $\alpha = 2$: s at most $\sim c2(\alpha,q1,q2)\ (\log n)^2$
  - $\alpha > 2$: s at least $\sim c3(\alpha,q1,q2)\ n^{(\alpha-2)/(\alpha-1)}$



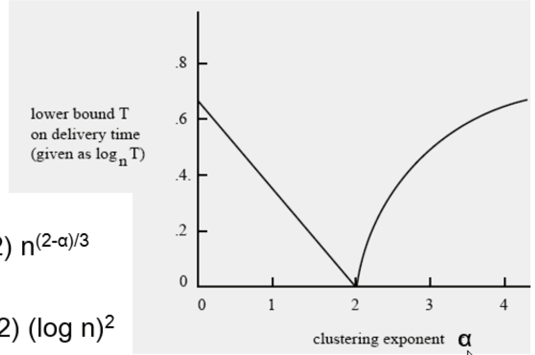
(closely after [3])

Kleinberg Model

- Now: Send message with local (decentralized) knowledge only

- Given: Decentralized greedy message delivery algorithm: measure number of expected delivery steps s:

  - $0 \le \alpha < 2$: s at least $\sim c1(\alpha,q1,q2)\ n^{(2-\alpha)/3}$
  - $\alpha = 2$: s at most $\sim c2(\alpha,q1,q2)\ (\log n)^2$
  - $\alpha > 2$: s at least $\sim c3(\alpha,q1,q2)\ n^{(\alpha-2)/(\alpha-1)}$



(closely after [3])

Kleinberg Model

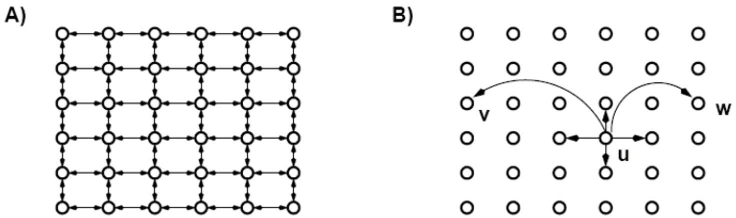
- Put nodes on a n x n grid. Distance: Manhattan:

$$r(i,j) = |x_i - x_j| + |y_i - y_j|$$

- Each node i: Connected to all nodes with $r(i,j) \le q1$ (regular local contacts)
- Each node i: Additional q2 other „long range" edges:
  Probability of edge to node j:

$$P(j) \sim r(i,j)^{-\alpha}$$

examples:



q1 = 1, q2 = 0

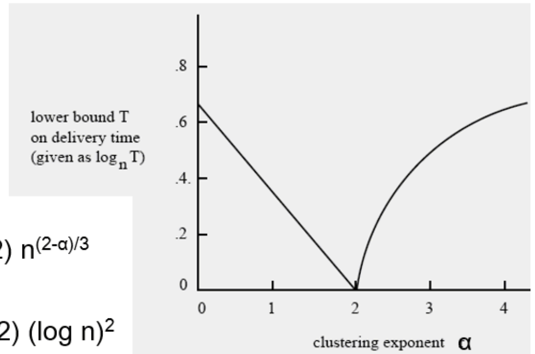
edges for one node u: (q1 = 1, q2 = 2) [3]

Kleinberg Model

- Now: Send message with local (decentralized) knowledge only

- Given: Decentralized greedy message delivery algorithm: measure number of expected delivery steps s:

  - $0 \le \alpha < 2$: s at least $\sim c1(\alpha,q1,q2)\ n^{(2-\alpha)/3}$
  - $\alpha = 2$: s at most $\sim c2(\alpha,q1,q2)\ (\log n)^2$
  - $\alpha > 2$ s at least $\sim c3(\alpha,q1,q2)\ n^{(\alpha-2)/(\alpha-1)}$

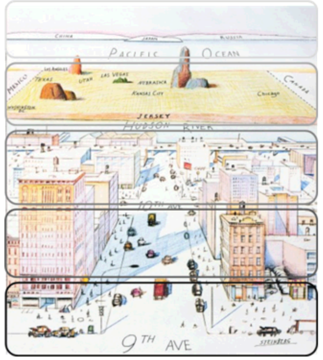


(closely after [3])

Kleinberg Model

- Effectively: Only a notion of distance (not neccess. spatial!) is necessary to route!
- Applications in P2P Systems (see [2])

- Explanation for the dependency of $\alpha=D$ on the grid‘s dimension D for efficient delivery:
  - start from node u
  - partition the set of other nodes into sets $A_0, A_1, A_2, \ldots, A_{\log n}$, where A_i has a distance to u between 2^i and 2^{i+1}
  - proven in [3]: only $\alpha=D$ ensures that the q2 „long-range“ contacts are evenly distributed over the A_i
  - for $\alpha>D$: bias towards smaller distances; for $\alpha<D$: bias towards larger distances

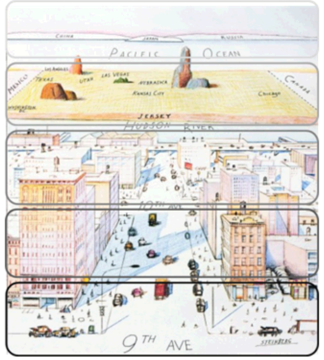


[5,6]

Kleinberg Model

- Effectively: Only a notion of distance (not neccess. spatial!) is necessary to route!
- Applications in P2P Systems (see [2])

- Explanation for the dependency of $\alpha=D$ on the grid‘s dimension D for efficient delivery:
  - start from node u
  - partition the set of other nodes into sets $A_0, A_1, A_2, \ldots, A_{\log n}$, where A_i has a distance to u between 2^i and 2^{i+1}
  - proven in [3]: only $\alpha=D$ ensures that the q2 „long-range“ contacts are evenly distributed over the A_i
  - for $\alpha>D$: bias towards smaller distances; for $\alpha<D$: bias towards larger distances

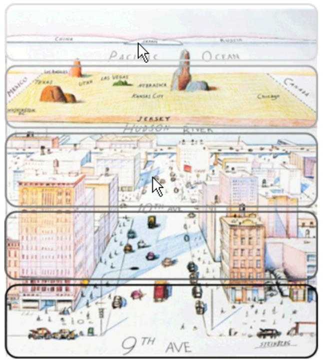


[5,6]

Kleinberg Model

- Effectively: Only a notion of distance (not neccess. spatial!) is necessary to route!
- Applications in P2P Systems (see [2])

- Explanation for the dependency of $\alpha=D$ on the grid‘s dimension D for efficient delivery:
  - start from node u
  - partition the set of other nodes into sets $A_0, A_1, A_2, \ldots, A_{\log n}$, where A_i has a distance to u between 2^i and 2^{i+1}
  - proven in [3]: only $\alpha=D$ ensures that the q2 „long-range“ contacts are evenly distributed over the A_i
  - for $\alpha>D$: bias towards smaller distances; for $\alpha<D$: bias towards larger distances

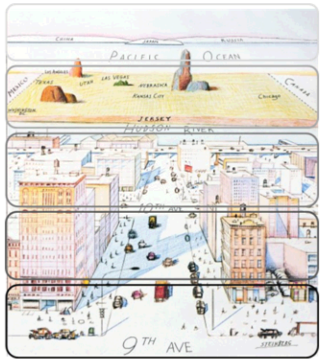


[5,6]

Kleinberg Model

- Effectively: Only a notion of distance (not neccess. spatial!) is necessary to route!
- Applications in P2P Systems (see [2])

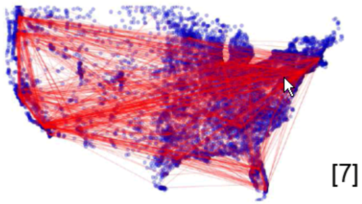
- Explanation for the dependency of $\alpha=D$ on the grid‘s dimension D for efficient delivery:
  - start from node u
  - partition the set of other nodes into sets $A_0, A_1, A_2, \ldots, A_{\log n}$, where A_i has a distance to u between 2^i and 2^{i+1}
  - proven in [3]: only $\alpha=D$ ensures that the q2 „long-range“ contacts are evenly distributed over the A_i
  - for $\alpha>D$: bias towards smaller distances; for $\alpha<D$: bias towards larger distances



[5,6]

Geographic Distance as Routing Metric

- In analysis of Milgram's experiment:
  - People early in the resp. path: often cite geographic proximity as main forwarding criterion;
  - People late in the path: chose similarity of occupation
- [6]: Study „Kleinberg-like" distributions / effects on real social NW:
  - LifeJournal.com : locate ~ 10^6 users (long/lat of their hometown)
  - Simulate Milgram on friendship NW with greedy decentralized forwarding and geographical proximity as criterion
  - Result: efficient routing is possible on average (see [6])

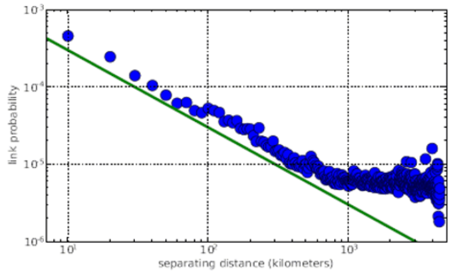


[7]

Fig. 2 The LiveJournal social network [32]. A dot is shown for each geographic location that was declared as the hometown of at least one of the ≈500,000 LiveJournal users whom we were able to locate at a longitude and latitude in the continental United States. A random 0.1% of the friendships in the network are overlaid on these locations.

Geographic Distance as Routing Metric

- But: Investigate Kleinberg's $\alpha \overset{!}{=} 2$ claim: result: probability of friendship as a function of distance reveals $\alpha \approx 1$



$$P(d) = \varepsilon + \Theta(1/d)$$

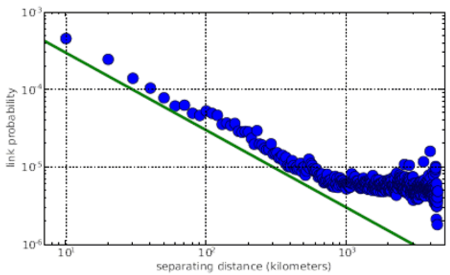
[7]

Fig. 3 The probability $P(d)$ of a friendship between two people in LiveJournal as a function of the geographic distance d between their declared hometowns [32]. Distances are rounded into 10-kilometer buckets. The solid line corresponds to $P(d) \propto 1/d$. Note that Theorem 1 requires $P(d) \propto 1/d^2$ for a network of people arranged in a regular 2-dimensional grid to be navigable.

- Although $\alpha \approx 1$ was found, systematic power law is surprising in "pure virtual" community as LiveJournal

Geographic Distance as Routing Metric

- But: Investigate Kleinberg's $\alpha \overset{!}{=} 2$ claim: result: probability of friendship as a function of distance reveals $\alpha \approx 1$



$$P(d) = \varepsilon + \Theta(1/d)$$

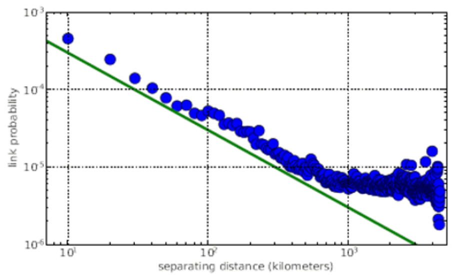
[7]

Fig. 3 The probability $P(d)$ of a friendship between two people in LiveJournal as a function of the geographic distance d between their declared hometowns [32]. Distances are rounded into 10-kilometer buckets. The solid line corresponds to $P(d) \propto 1/d$. Note that Theorem 1 requires $P(d) \propto 1/d^2$ for a network of people arranged in a regular 2-dimensional grid to be navigable.

- Although $\alpha \approx 1$ was found, systematic power law is surprising in "pure virtual" community as LiveJournal

Geographic Distance as Routing Metric

- But: Investigate Kleinberg's $\alpha \overset{!}{=} 2$ claim: result: probability of friendship as a function of distance reveals $\alpha \approx 1$



$$P(d) = \varepsilon + \Theta(1/d)$$

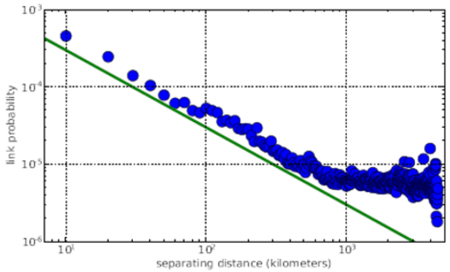
[7]

Fig. 3 The probability $P(d)$ of a friendship between two people in LiveJournal as a function of the geographic distance d between their declared hometowns [32]. Distances are rounded into 10-kilometer buckets. The solid line corresponds to $P(d) \propto 1/d$. Note that Theorem 1 requires $P(d) \propto 1/d^2$ for a network of people arranged in a regular 2-dimensional grid to be navigable.

- Although $\alpha \approx 1$ was found, systematic power law is surprising in "pure virtual" community as LiveJournal

Geographic Distance as Routing Metric

- But: Investigate Kleinberg's $\alpha \overset{!}{=} 2$ claim: result: probability of friendship as a function of distance reveals $\alpha \approx 1$



$$P(d) = \varepsilon + \Theta(1/d)$$

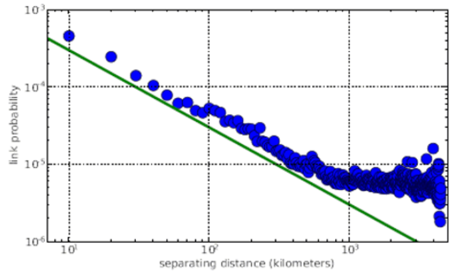
[7]

Fig. 3 The probability $P(d)$ of a friendship between two people in LiveJournal as a function of the geographic distance d between their declared hometowns [32]. Distances are rounded into 10-kilometer buckets. The solid line corresponds to $P(d) \propto 1/d$. Note that Theorem 1 requires $P(d) \propto 1/d^2$ for a network of people arranged in a regular 2-dimensional grid to be navigable.

- Although $\alpha \approx 1$ was found, systematic power law is surprising in "pure virtual" community as LiveJournal

Geographic Distance as Routing Metric

- But: Investigate Kleinberg's $\alpha \overset{!}{=} 2$ claim: result: probability of friendship as a function of distance reveals $\alpha \approx 1$



$$P(d) = \varepsilon + \Theta(1/d)$$

[7]

Fig. 3 The probability $P(d)$ of a friendship between two people in LiveJournal as a function of the geographic distance d between their declared hometowns [32]. Distances are rounded into 10-kilometer buckets. The solid line corresponds to $P(d) \propto 1/d$. Note that Theorem 1 requires $P(d) \propto 1/d^2$ for a network of people arranged in a regular 2-dimensional grid to be navigable.

- Although $\alpha \approx 1$ was found, systematic power law is surprising in "pure virtual" community as LiveJournal

Geographic Distance as Routing Metric

- Other studies: also confirm efficient decentralized routing with geographic proximity as criterion is possible AND $\alpha \approx 1$
- Explanation for this "contradiction": geographic density of people is not uniform (cp. urban vs. mid west) as assumed in the Kleinberg grid.
- Model: Rank-based friendship: modify Kleinberg (D=k):
  - at each grid point more than one person
  - p(v) for long range contact v starting at node u: $p_u(v) \sim 1 / rank_u(v)$ where rank(u,v) = number of people who live at least as close to u as v does. (→ u sorts candidates v according to distance)
  - (if population density is **uniform**: $rank_u(v) \sim \theta(d(u,v)^k)$ → Kleinberg's claim is fulfilled)

Geographic Distance as Routing Metric

- Other studies: also confirm efficient decentralized routing with geographic proximity as criterion is possible AND $\alpha \approx 1$
- Explanation for this "contradiction": geographic density of people is not uniform (cp. urban vs. mid west) as assumed in the Kleinberg grid.
- Model: Rank-based friendship: modify Kleinberg (D=k):
  - at each grid point more than one person
  - p(v) for long range contact v starting at node u: $p_u(v) \sim 1 / rank_u(v)$ where rank(u,v) = number of people who live at least as close to u as v does. (→ u sorts candidates v according to distance)
  - (if population density is **uniform**: $rank_u(v) \sim \theta(d(u,v)^k)$ → Kleinberg's claim is fulfilled)

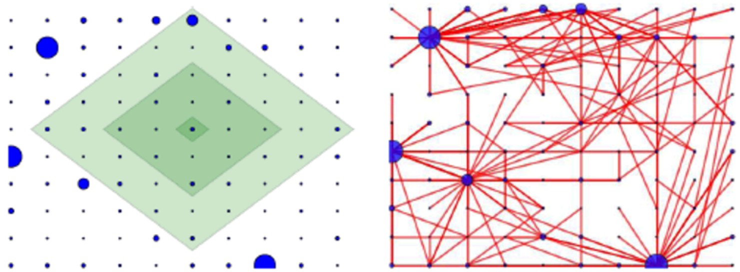
Geographic Distance as Routing Metric

- Other studies: also confirm efficient decentralized routing with geographic proximity as criterion is possible AND $\alpha \approx 1$
- Explanation for this “contradiction”: geographic density of people is not uniform (cp. urban vs. mid west) as assumed in the Kleinberg grid.

- Model: Rank-based friendship: modify Kleinberg (D=k):
  - at each grid point more than one person
  - p(v) for long range contact v starting at node u: $p_u(v) \sim 1 / rank_u(v)$ where rank(u,v) = number of people who live at least as close to u as v does. (→ u sorts candidates v according to distance)
  - (if population density is **uniform**: $rank_u(v) \sim \theta(d(u,v)^k)$ → Kleinberg‘s claim is fulfilled)

Geographic Distance as Routing Metric

- Finding: Even for non-uniform geographic distribution of people: efficient routing possible (choose pairs randomly):
  - For q1= 2k and q2 = 1 : $T \sim O((\log n)^3)$



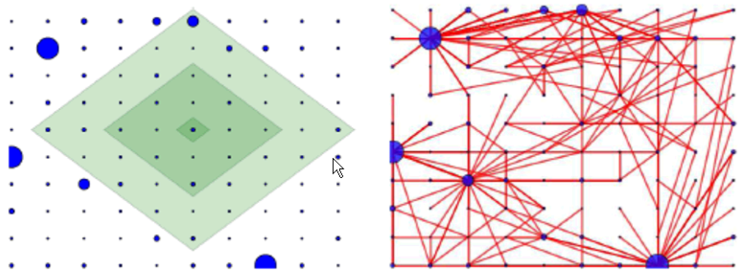
(a) Concentric balls around a city *C*, where each ball’s population increases by a factor of four. A resident of *C* choosing a rank-based friend is four times more likely to choose a friend at the boundary of one ball than a friend at the boundary of the next-larger ball.

(b) A rank-based social network generated from this population distribution. For visual simplicity, edges are depicted as connecting cities; the complete image would show each edge connecting one resident from each of its endpoint cities.

[7]

Geographic Distance as Routing Metric

- Finding: Even for non-uniform geographic distribution of people: efficient routing possible (choose pairs randomly):
  - For q1= 2k and q2 = 1 : $T \sim O((\log n)^3)$



(a) Concentric balls around a city *C*, where each ball’s population increases by a factor of four. A resident of *C* choosing a rank-based friend is four times more likely to choose a friend at the boundary of one ball than a friend at the boundary of the next-larger ball.

(b) A rank-based social network generated from this population distribution. For visual simplicity, edges are depicted as connecting cities; the complete image would show each edge connecting one resident from each of its endpoint cities.

[7]

Geographic Distance as Routing Metric

- Kleinberg’s version: random expectation is only taken over random construction of network. routing is regarded to any (fixed) target node

Rank-Based version (here): random expectation is taken over random construction of network AND random choice of target node.

--> In Kleinberg’s version each node is efficiently reachable; Here: node-pairs are chosen uniformly randomly → connections from dense grid points to other dense grid points are more likely. However: here: persons in thinly populated areas might be hard to route to.

Geographic Distance as Routing Metric

- Kleinberg's version: random expectation is only taken over random construction of network. routing is regarded to any (fixed) target node

Rank-Based version (here): random expectation is taken over random construction of network AND random choice of target node.

--> In Kleinberg's version each node is efficiently reachable; Here: node-pairs are chosen uniformly randomly → connections from dense grid points to other dense grid points are more likely. However: here: persons in thinly populated areas might be hard to route to.

Geographic Distance as Routing Metric

- Kleinberg's version: random expectation is only taken over random construction of network. routing is regarded to any (fixed) target node

Rank-Based version (here): random expectation is taken over random construction of network AND random choice of target node.

--> In Kleinberg's version each node is efficiently reachable; Here: node-pairs are chosen uniformly randomly → connections from dense grid points to other dense grid points are more likely. However: here: persons in thinly populated areas might be hard to route to.

Off-Grid / Other Metrics

- Grid and concentration on geo-proximity alone: not realistic.
- Example: occupation: given taxonomy of occupations: similarity measure: determine tree height of least common ancestor (lca) of u and v
- Kleinberg: if tree is a regular b-ary tree and long range probability goes as $\Pr[u \rightarrow v] \propto b^{-\beta \cdot \text{lca}(u,v)}$ then efficient routing only for β=1
- Generalization: tree → graph : Pr[u→v] ~ 1 / f (path-distance(u,v))

Off-Grid / Other Metrics

- Grid and concentration on geo-proximity alone: not realistic.
- Example: occupation: given taxonomy of occupations: similarity measure: determine tree height of least common ancestor (lca) of u and v
- Kleinberg: if tree is a regular b-ary tree and long range probability goes as $\Pr[u \rightarrow v] \propto b^{-\beta \cdot \text{lca}(u,v)}$ then efficient routing only for β=1
- Generalization: tree → graph : Pr[u→v] ~ 1 / f (path-distance(u,v))

Off-Grid / Other Metrics

- Grid and concentration on geo-proximity alone: not realistic.
- Example: occupation: given taxonomy of occupations: similarity measure: determine tree height of least common ancestor (lca) of u and v
- Kleinberg: if tree is a regular b-ary tree and long range probability goes as $\Pr[u \rightarrow v] \propto b^{-\beta \cdot \mathrm{lca}(u,v)}$ then efficient routing only for β=1
- Generalization: tree → graph : Pr[u→v] ~ 1 / f (path-distance(u,v))

Off-Grid / Other Metrics

- Grid and concentration on geo-proximity alone: not realistic.
- Example: occupation: given taxonomy of occupations: similarity measure: determine tree height of least common ancestor (lca) of u and v
- Kleinberg: if tree is a regular b-ary tree and long range probability goes as $\Pr[u \rightarrow v] \propto b^{-\beta \cdot \mathrm{lca}(u,v)}$ then efficient routing only for β=1
- Generalization: tree → graph : Pr[u→v] ~ 1 / f (path-distance(u,v))

Off-Grid / Other Metrics

- Kleinberg: Group Model: Partition set of actors into groups: Pr[u→v] ~ [1 / size of smallest common group of u and v] → navigateable with decentralized greedy approach if:
  - Groups form a hierarchy (small contained in large) → „narrowing in“ possible
  - Group-sizes satisfy certain bounds → non-zero „escaping“ probability from a small group
- Advantage: Groups can be formed according to different criteria simultaneously
- Strictly spoken: Even in the 2-dim grid: we have a combination of two elements (long and lat)

Off-Grid / Other Metrics

- Kleinberg: Group Model: Partition set of actors into groups: Pr[u→v] ~ [1 / size of smallest common group of u and v] → navigateable with decentralized greedy approach if:
  - Groups form a hierarchy (small contained in large) → „narrowing in“ possible
  - Group-sizes satisfy certain bounds → non-zero „escaping“ probability from a small group
- Advantage: Groups can be formed according to different criteria simultaneously
- Strictly spoken: Even in the 2-dim grid: we have a combination of two elements (long and lat)

Off-Grid / Other Metrics

- Kleinberg: Group Model: Partition set of actors into groups: Pr[u→v] ~ [1 / size of smallest common group of u and v] → navigateable with decentralized greedy approach if:
  - Groups form a hierarchy (small contained in large) → „narrowing in“ possible
  - Group-sizes satisfy certain bounds → non-zero „escaping“ probability from a small group
- Advantage: Groups can be formed according to different criteria simultaneously
- Strictly spoken: Even in the 2-dim grid: we have a combination of two elements (long and lat)

Off-Grid / Other Metrics

- Kleinberg: Group Model: Partition set of actors into groups: Pr[u→v] ~ [1 / size of smallest common group of u and v] → navigateable with decentralized greedy approach if:
  - Groups form a hierarchy (small contained in large) → „narrowing in“ possible
  - Group-sizes satisfy certain bounds → non-zero „escaping“ probability from a small group
- Advantage: Groups can be formed according to different criteria simultaneously
- Strictly spoken: Even in the 2-dim grid: we have a combination of two elements (long and lat)

Off-Grid / Other Metrics

- Kleinberg: Group Model: Partition set of actors into groups: Pr[u→v] ~ [1 / size of smallest common group of u and v] → navigateable with decentralized greedy approach if:
  - Groups form a hierarchy (small contained in large) → „narrowing in“ possible
  - Group-sizes satisfy certain bounds → non-zero „escaping“ probability from a small group
- Advantage: Groups can be formed according to different criteria simultaneously
- Strictly spoken: Even in the 2-dim grid: we have a combination of two elements (long and lat)

Off-Grid / Other Metrics

- Watts et al: extend the tree model: now: multiple hierarchies: if $lca_i(u,v)$ is the least common ancestor in the ith hierarchy → set

$$d(u,v) := \min_i \mathrm{lca}_i(u,v)$$

  - for the multi-hierarchy: a wider range of exponents β (other than β =1) yields navigatable networks (with polylogarithmic delivery times)
- Other variants: include semi-local knowledge into routing algorithm: promising: know the number of friends of a user (models up to here assume an almost constant degree; now: exploit power law)
- Other variants: consider degree distributions (e.g. power law) other than constant degree (as before) (people with more friends can act as „hubs“)

Off-Grid / Other Metrics

- Watts et al: extend the tree model: now: multiple hierarchies: if $\mathsf{lca}_i(u,v)$ is the least common ancestor in the ith hierarchy → set

$$d(u,v) := \min_i \mathsf{lca}_i(u,v)$$

  - for the multi-hierarchy: a wider range of exponents β (other than β =1) yields navigatable networks (with polylogarithmic delivery times)

- Other variants: include semi-local knowledge into routing algorithm: promising: know the number of friends of a user (models up to here assume an almost constant degree; now: exploit power law)

- Other variants: consider degree distributions (e.g. power law) other than constant degree (as before) (people with more friends can act as „hubs“)

Off-Grid / Other Metrics

- Watts et al: extend the tree model: now: multiple hierarchies: if $\mathsf{lca}_i(u,v)$ is the least common ancestor in the ith hierarchy → set

$$d(u,v) := \min_i \mathsf{lca}_i(u,v)$$

  - for the multi-hierarchy: a wider range of exponents β (other than β =1) yields navigatable networks (with polylogarithmic delivery times)

- Other variants: include semi-local knowledge into routing algorithm: promising: know the number of friends of a user (models up to here assume an almost constant degree; now: exploit power law)

- Other variants: consider degree distributions (e.g. power law) other than constant degree (as before) (people with more friends can act as „hubs“)

Off-Grid / Other Metrics

- Watts et al: extend the tree model: now: multiple hierarchies: if $\mathsf{lca}_i(u,v)$ is the least common ancestor in the ith hierarchy → set

$$d(u,v) := \min_i \mathsf{lca}_i(u,v)$$

  - for the multi-hierarchy: a wider range of exponents β (other than β =1) yields navigatable networks (with polylogarithmic delivery times)

- Other variants: include semi-local knowledge into routing algorithm: promising: know the number of friends of a user (models up to here assume an almost constant degree; now: exploit power law)

- Other variants: consider degree distributions (e.g. power law) other than constant degree (as before) (people with more friends can act as „hubs“)

Off-Grid / Other Metrics

- Watts et al: extend the tree model: now: multiple hierarchies: if $\mathsf{lca}_i(u,v)$ is the least common ancestor in the ith hierarchy → set

$$d(u,v) := \min_i \mathsf{lca}_i(u,v)$$

  - for the multi-hierarchy: a wider range of exponents β (other than β =1) yields navigatable networks (with polylogarithmic delivery times)

- Other variants: include semi-local knowledge into routing algorithm: promising: know the number of friends of a user (models up to here assume an almost constant degree; now: exploit power law)

- Other variants: consider degree distributions (e.g. power law) other than constant degree (as before) (people with more friends can act as „hubs“)

Off-Grid / Other Metrics

- Example of combining degree based heuristics with distance based heuristics: expected value navigation EVN: choose node v, whose probability of a direct link to target node t is maximal
- If p_{vt} is the probability that v connects directly to t (e.g. insert Kleinberg's long-range distance based probability $p_{vt} = d(v,t)^{-\alpha}$ here) and degree of v is $\delta(v)$ we then have
  - probability that at least one of v's friends will provide the direct link to t is then $1 - (1 - p_{vt})^{\delta(v)}$
  - → choose v that maximizes this probability

Off-Grid / Other Metrics

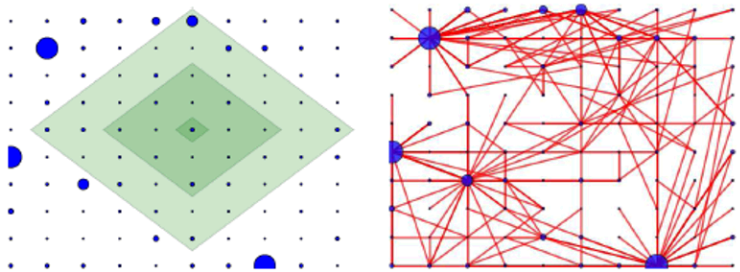
- Example of combining degree based heuristics with distance based heuristics: expected value navigation EVN: choose node v, whose probability of a direct link to target node t is maximal
- If p_{vt} is the probability that v connects directly to t (e.g. insert Kleinberg's long-range distance based probability $p_{vt} = d(v,t)^{-\alpha}$ here) and degree of v is $\delta(v)$ we then have
  - probability that at least one of v's friends will provide the direct link to t is then $1 - (1 - p_{vt})^{\delta(v)}$
  - → choose v that maximizes this probability

Off-Grid / Other Metrics

- Example of combining degree based heuristics with distance based heuristics: expected value navigation EVN: choose node v, whose probability of a direct link to target node t is maximal
- If p_{vt} is the probability that v connects directly to t (e.g. insert Kleinberg's long-range distance based probability $p_{vt} = d(v,t)^{-\alpha}$ here) and degree of v is $\delta(v)$ we then have
  - probability that at least one of v's friends will provide the direct link to t is then $1 - (1 - p_{vt})^{\delta(v)}$
  - → choose v that maximizes this probability

Geographic Distance as Routing Metric

- Finding: Even for non-uniform geographic distribution of people: efficient routing possible (choose pairs randomly):
  - For $q1 = 2k$ and $q2 = 1$: $T \sim O((\log n)^3)$



(a) Concentric balls around a city C, where each ball's population increases by a factor of four. A resident of C choosing a rank-based friend is four times more likely to choose a friend at the boundary of one ball than a friend at the boundary of the next-larger ball.

(b) A rank-based social network generated from this population distribution. For visual simplicity, edges are depicted as connecting cities; the complete image would show each edge connecting one resident from each of its endpoint cities.

[7]

Geographic Distance as Routing Metric

- But: Investigate Kleinberg's $\alpha \overset{!}{=} 2$ claim: result: probability of friendship as a function of distance reveals $\alpha \approx 1$

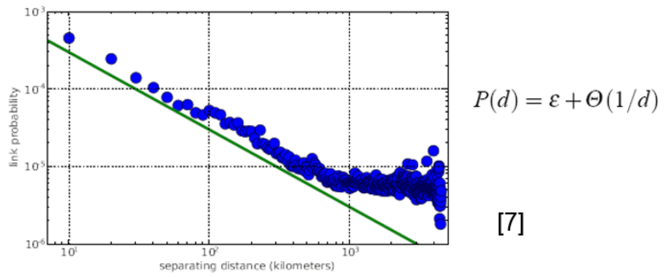
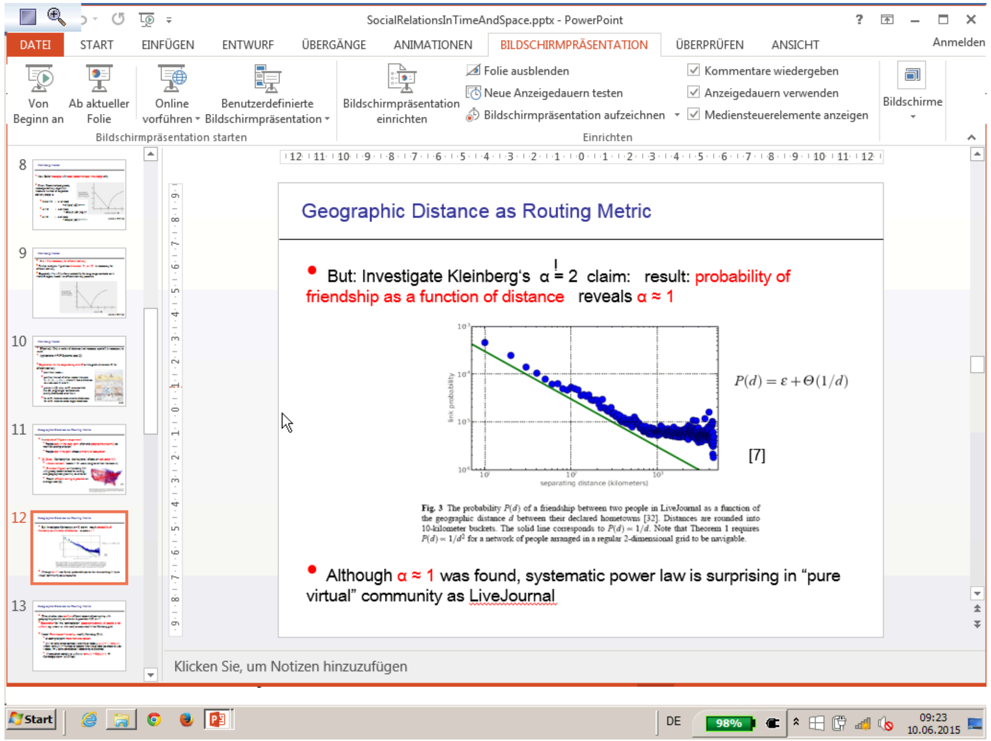


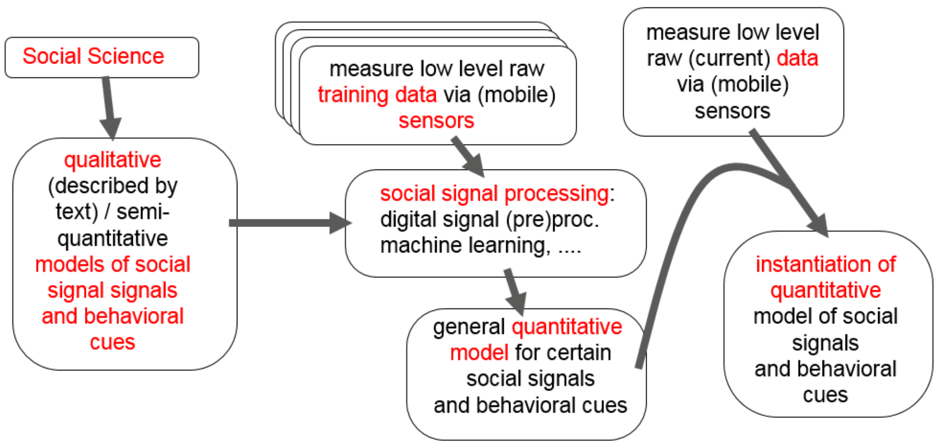
Fig. 3 The probability $P(d)$ of a friendship between two people in LiveJournal as a function of the geographic distance d between their declared hometowns [32]. Distances are rounded into 10-kilometer buckets. The solid line corresponds to $P(d) \propto 1/d$. Note that Theorem 1 requires $P(d) \propto 1/d^2$ for a network of people arranged in a regular 2-dimensional grid to be navigable.

- Although $\alpha \approx 1$ was found, systematic power law is surprising in "pure virtual" community as LiveJournal



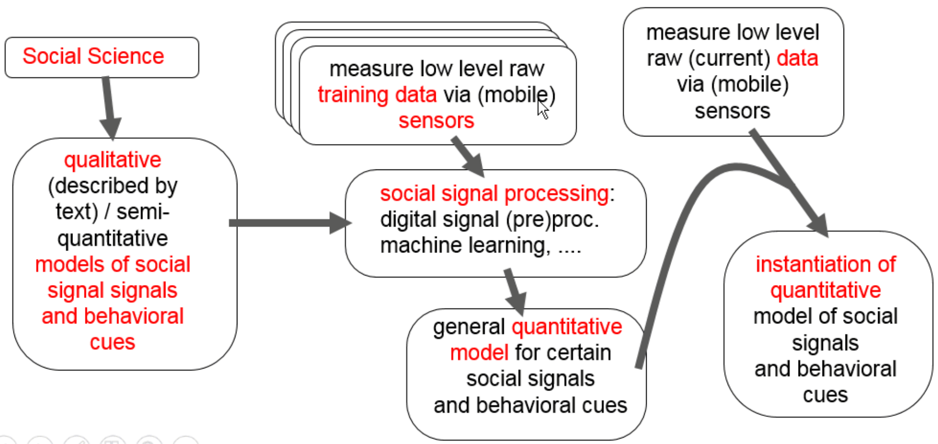
Social Signal Processing

- Up to now: most emphasis on long term social context (e.g. networks of friendships)
- Now: short term social context: social behavior on small spatial and temporal scales: detection + applications



Social Signal Processing

- Up to now: most emphasis on long term social context (e.g. networks of friendships)
- Now: short term social context: social behavior on small spatial and temporal scales: detection + applications



Social Signal Processing

- General field: Human behavior modeling
  → lots of „options" for individual
  especially if individual's behavioral entropy is high
  → individual behavior models often need (lots of)
  individual training data

- Basic social behavior
  follows strict, nearly uniform rules (within certain cultural frames (groups))
  → social behavior models can exploit these uniform rules
  → models may not have to be specially trained for individuals