Referral Systems:
Applications in Knowledge Management; Emergent Properties

Munindar P. Singh
(joint work with Bin Yu, pInar Yolum (mainly), Yathi Udupi)
Department of Computer Science
North Carolina State University

Outline

- Motivation and Framework
- Making Referral Systems Useful
- Understanding Referral Systems
  - Authoritativeness
  - Structure
  - Evolution
- Directions
- Backup
  - Clustering
  - Power-law networks

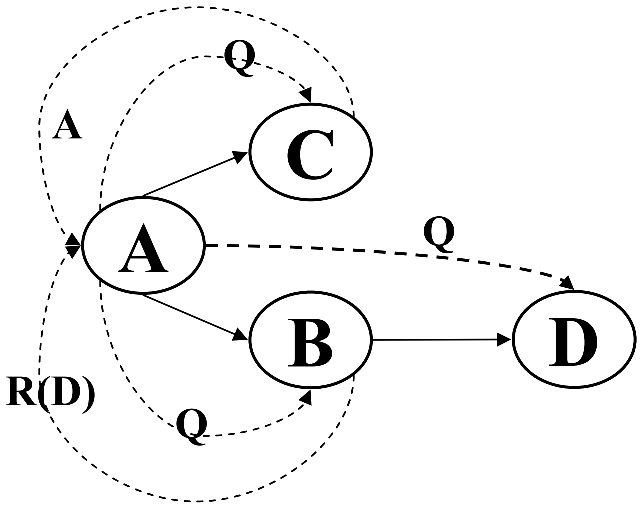
Knowledge Management

Management of knowledge at the individual and enterprise levels

- The traditional approach is *artifact centric:* focuses on documents mainly
- Major shortcoming: most valuable knowledge is not in artifacts
  - *Ownership:* opposed to individual interests
  - *Lack of context:* where applied
  - *Violation of privacy:* how much would you reveal
  - *Need know-how:* not just know-that
- Instead find people who know

Abstraction: Referrals for Selection

How can we find a business partner in a purely distributed system?



- An agent represents a principal offering or searching for services
- An agent generates a query for a service; sends it to its *neighbors*
- Each neighbor may provide the service or refer to other agents (based on its *referral policies*)
- Each agent models the *expertise* (quality of a service) and *sociability* (quality of the referrals) of its acquaintances
- Based on these models, each agent can change its set of neighbors (using its *neighbor selection policy*): locally, autonomously
- *Social network:* as induced by the neighborhood relation

Why a Decentralized Approach?

Problems with central authorities (e.g., Verisign) or reputation systems (e.g., eBay)

- *Context and understanding:* The contexts of usage may differ
- *Empirical basis:* Best to trust experience
  - Did Verisign itself buy DVDs from Amazon?
- *Privacy:* Raters may not want to reveal true ratings in public
- *Trust:* Users of ratings don't necessarily know where the ratings come from

Motivation

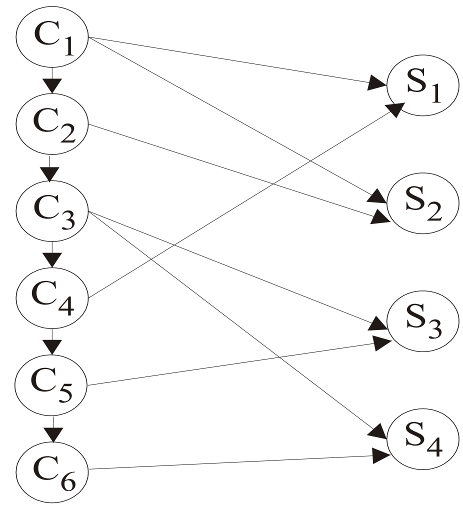
- Referrals for service selection
  - Follow referrals from trusted parties
  - Self-organize based on previous interactions
- Web structure
  - Properties of its snapshot
  - Stochastic models for approximating in-degree distributions
  - Hyperlinks are assumed to be endorsements
  - Local interactions are not captured

Referral process is explicit; emergent structure is not

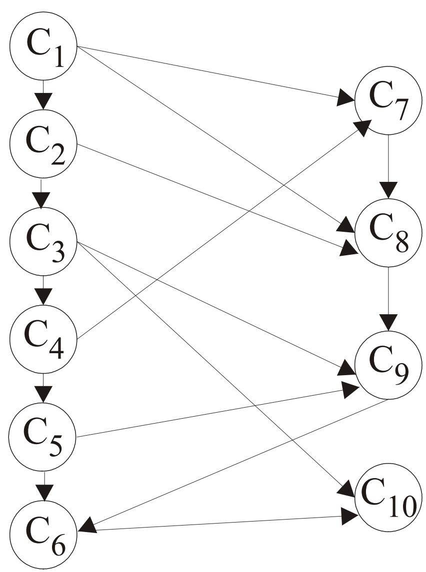
Emergent structure is explicit; underlying process is not

Application Domains

Commerce:



- Distinct service producers and consumers
- Producers have expertise, consumers have sociability
- Answers are easy to evaluate
- Expertise of consumers does not increase



Knowledge Management:

- All agents can be producers and consumers
- Answers are harder to evaluate
- Expertise of consumers may increase (expertise of the producers can be cached by others)

MARS: MultiAgent Referral System

Prototype system for helping people participate in a referral network

- Practical challenges:
  - UI: use an IM client
  - Communication: use an IM server (Jabber)
  - Bootstrap: Infer people's expertise and (initial) neighbors: mine email
- Research challenges
  - How to evaluate convincingly?

Developed over several years by
Bin Yu
Wentao Mo
Paul Palathingal
Subhayu Chatterjee

Representations: 1

- For simplicity, use vector representations for queries and knowledge
  - Assume dimensions; supply values
  - [spicy, timely, tasty, authentic, healthy]: [0.8,0.7,0.9,0.8,0.1]
- Easy approach conceptually
  - Common in text retrieval
  - Supports caching results
  - But has well-known limitations

Vector Space Model Originated in the 1960s
Still used in text retrieval

Applied by Yu & Singh; Yolum & Singh; Udupi, Yolum, & Singh

Representations: 2

- The meanings of the dimensions are not standard
- Ontology (loosely, conceptual model) for *qualities* of service
  - Common QoS: price, availability
  - Domain-specific QoS: spiciness
  - Idiosyncratic QoS: enjoyment
- How to handle preferences
  - Decision theory

Maximilien & Singh;

Maximilien developed a practical framework for QoS in Web services

QoS framework works as a reputation system; not yet combined into a referral system

Propagation of Trust

- Referrals support trust management
  - Provide a basis for finding *witnesses*, who can offer evidence (pro or con) about a third party
  - Provide a basis for rating such witnesses
  - Support adapting to select the more promising witnesses and avoid those who are deceptive

Yu & Singh: Applies Dempster-Shafer theory of evidence and weighted majority learning

Analysis

- Not just develop a system and hope it works, but understand its functioning to:
  - Improve its effectiveness in important settings
  - Find new uses for it
  - Study general questions of the consequences of decentralization and emergence

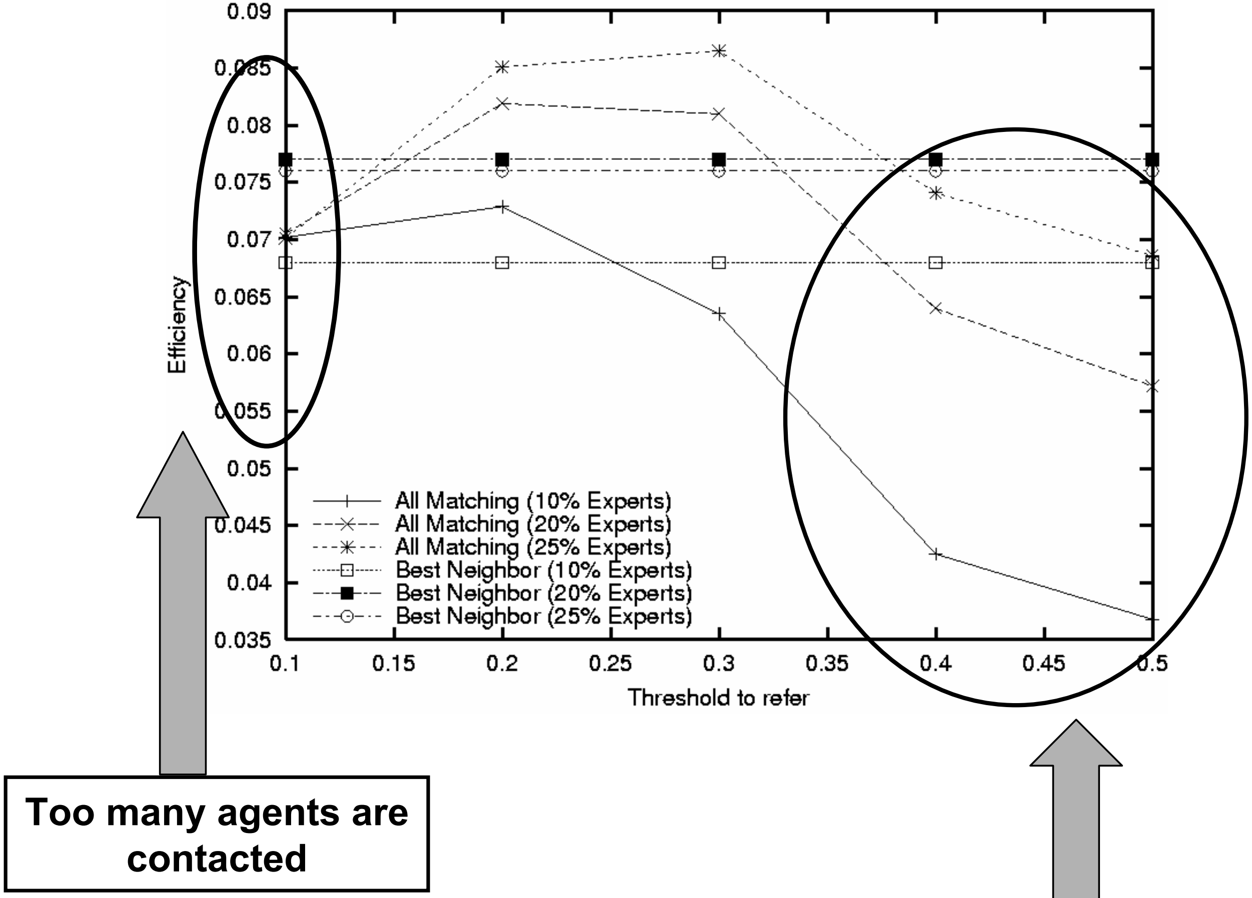
The completed work has mostly had an empirical flavor

Theoretical aspects would be great topics for further research

Referral Policies

Refer all neighbors:	Does not consider which neighbors would be more likely to answer (similar to Gnutella)
Refer all matching neighbors:	Refer those neighbors with "sufficient" expertise
Refer best neighbor:	Refer the most capable neighbor. Guarantees that at least one neighbor is referred

Efficiency of Referral Policies



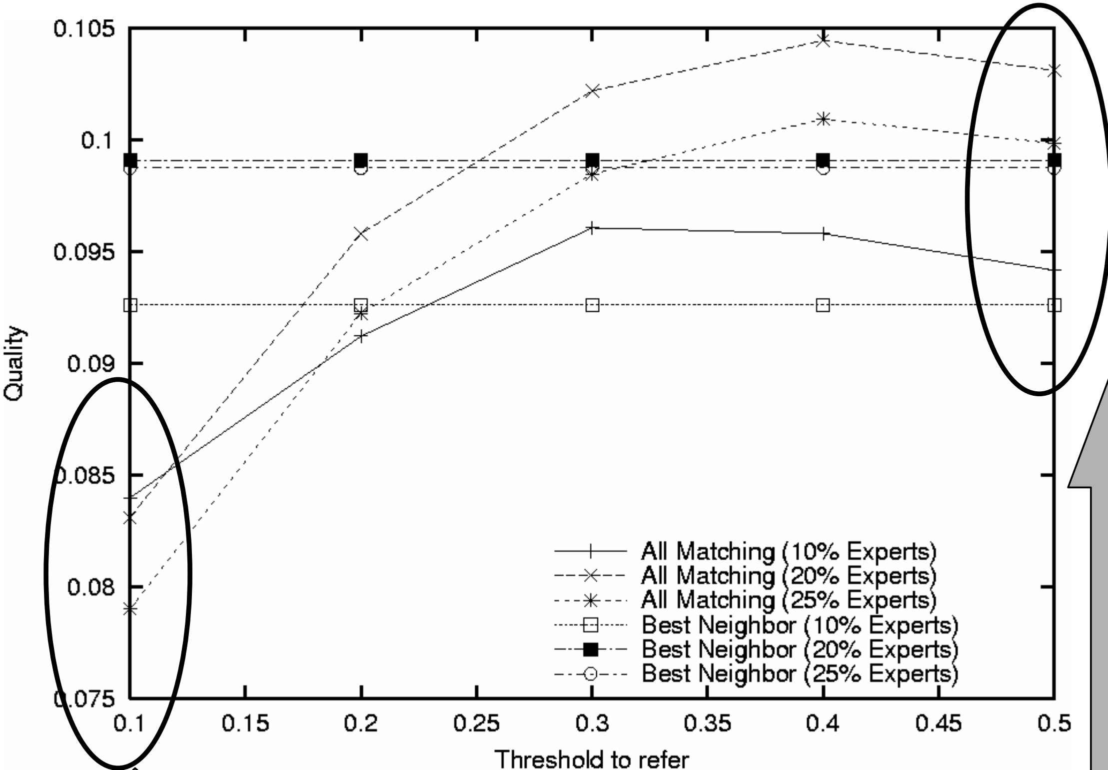
Policies:

Refer All
Refer All Matching
Refer Best

Efficiency =

$$\frac{\text{\# of good answers}}{\text{\# of contacted agents}}$$

Effectiveness of Referral Policies



Low quality even though answers are found

Low efficiency but high quality

For each agent A

- Rank all others (e.g., B) using $\frac{I_A \otimes E_B}{path(A,B)}$
- Take the nth highest
- n is twice the neighbor set size (8, 16)

Authorities

- Link analysis to find authorities from Web crawls
- PageRank: Pages pointed to by authorities are also authoritative

$$P(i) = d \sum_{j \in K_i} \frac{P(j)}{|N_j|} + (1 - d)$$

- Factors that influence the emergence of authorities

P(i): PageRank of i
N(j): Neighbors of j
K(i): Pages that point to page I
d: Damping factor

Referrals and Authorities

- Web search engines
  - Mostly crawl static pages
  - Interpret each URL as an endorsement
  - Mine centrally to decide where to direct searches by *all* users
- Referral systems
  - A decentralized agent
    - Obtains dynamic (custom) information
    - Knows if it is an endorsement
    - Decides how to use it for its user
    - Reveals appropriate information to others
  - Mining is optional, after the fact, for study and tuning

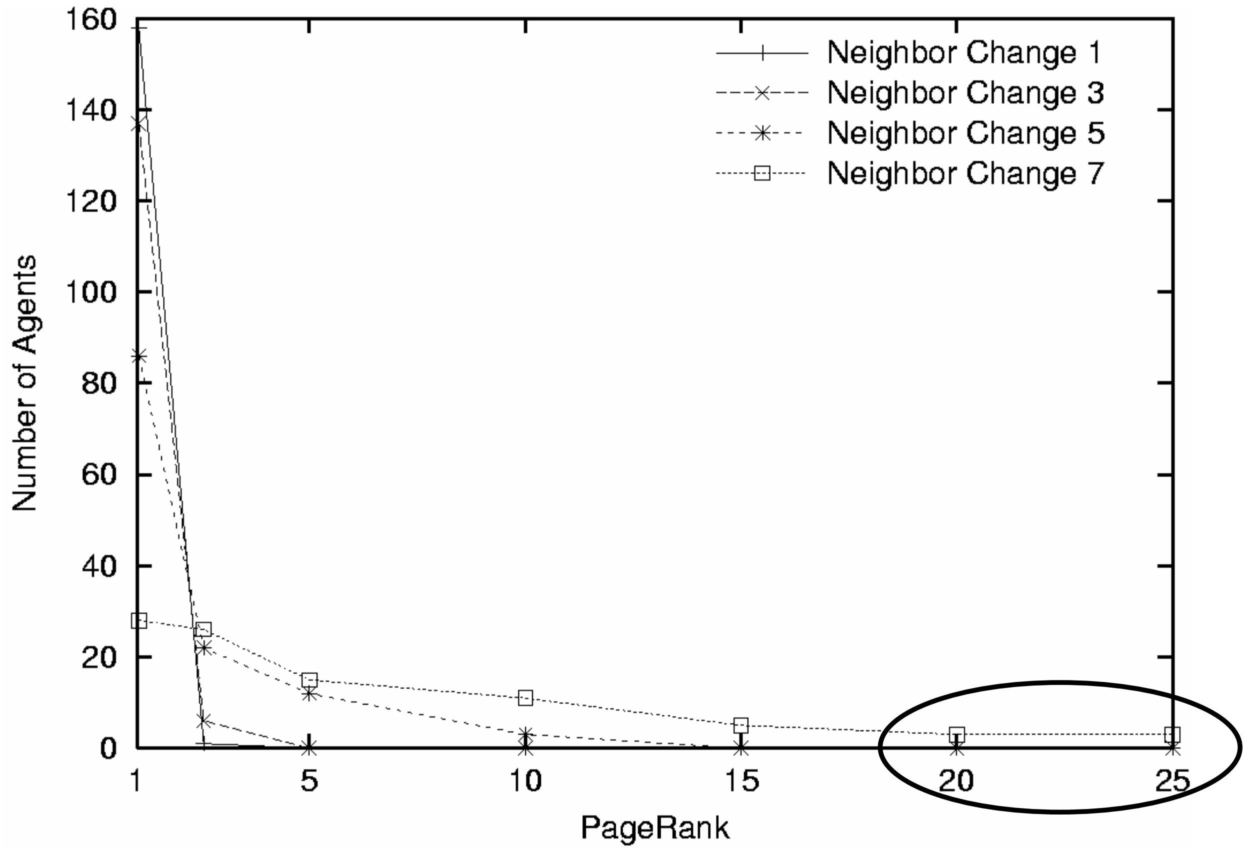
In referral systems, mining is used as a research tool

Cannot centrally crawl a referral system in practice

Exposing mined results may violate privacy

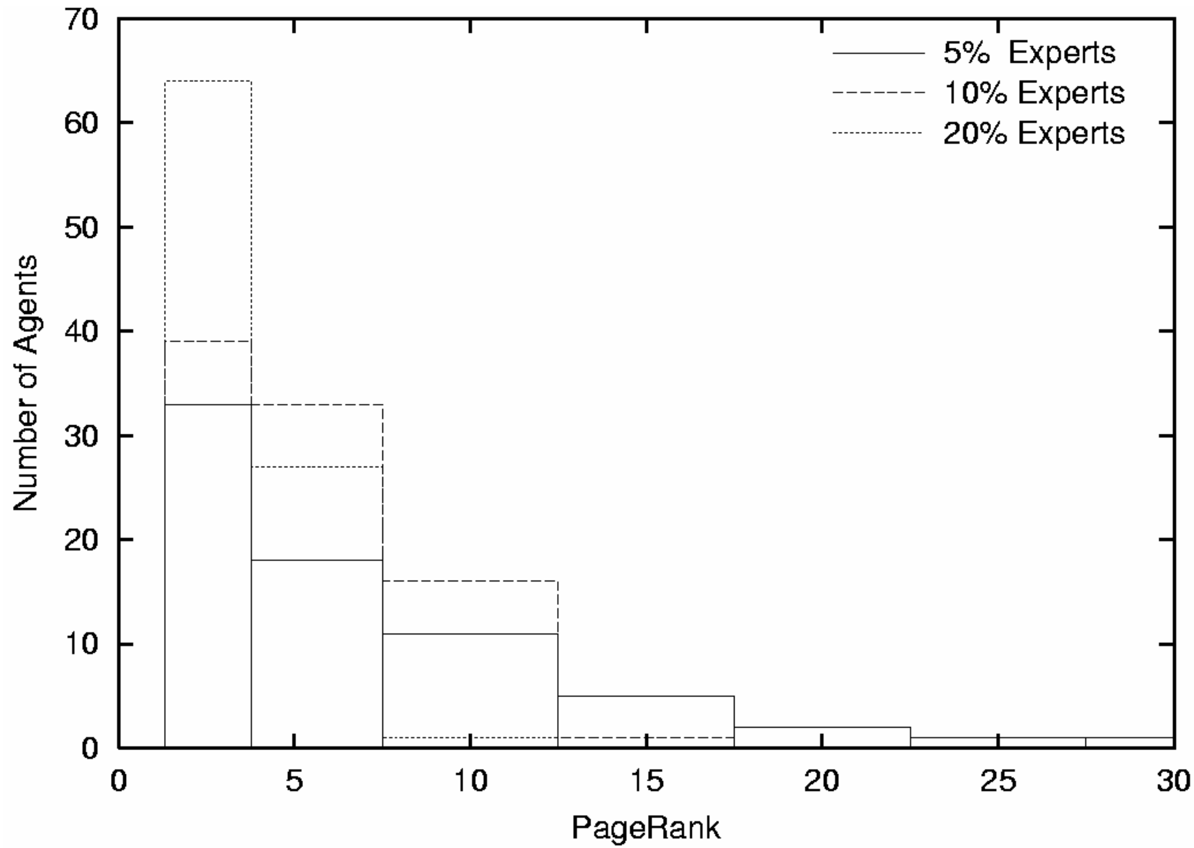
Yolum & Singh

Emergence of Authorities through Adaptation



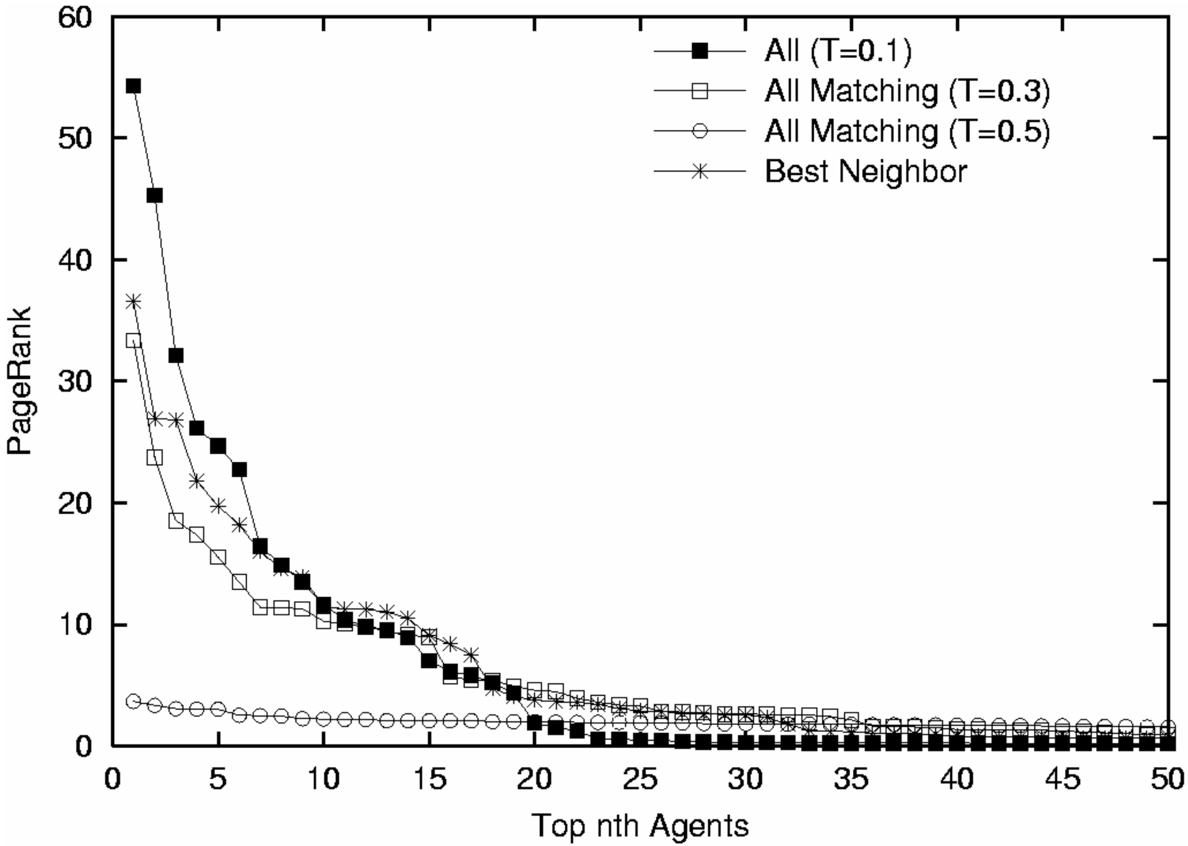
Authorities emerge as agents change neighbors

Authoritativeness & Number of Experts



When the population has fewer experts, the authoritativeness of the experts is higher

Effect of Referral Policies



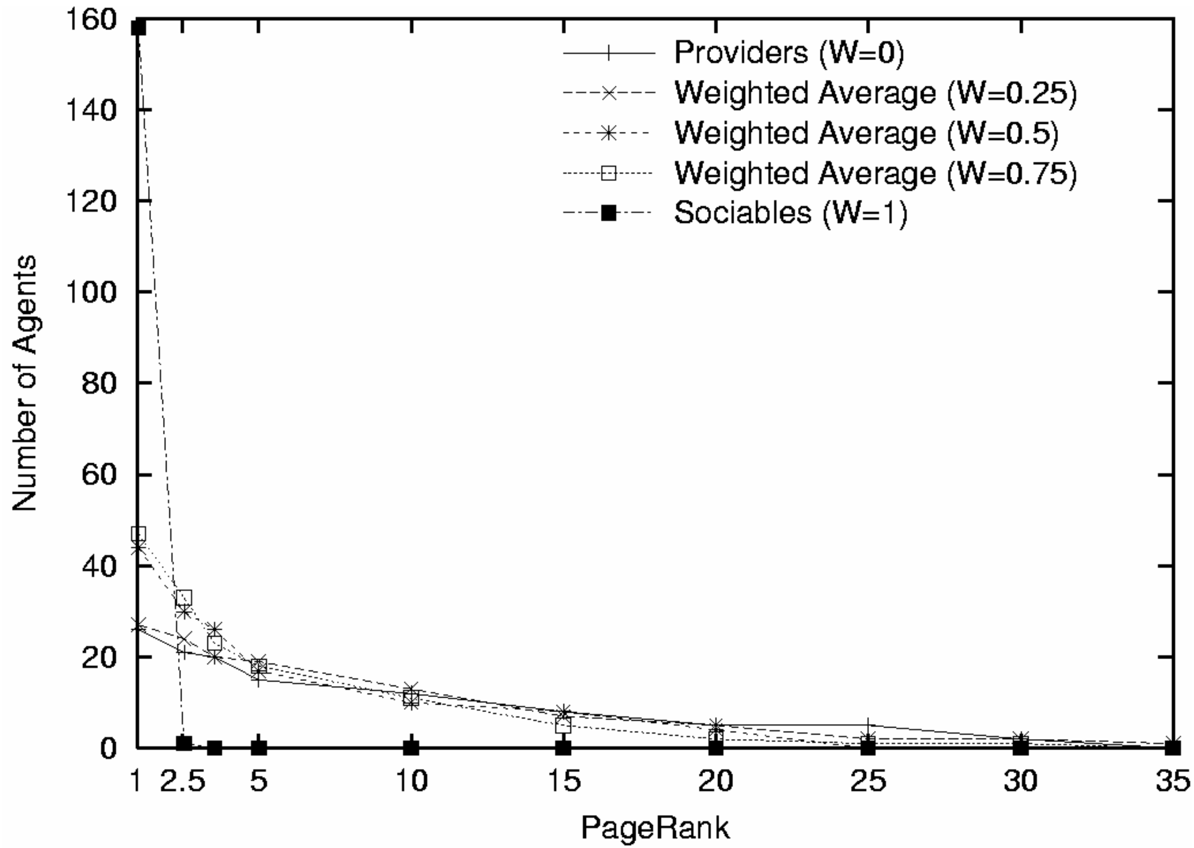
When more referrals are exchanged, the authorities obtain higher PageRank (i.e., their authoritativeness is greater)

Neighbor Selection Policies

How do the agents choose their neighbors?

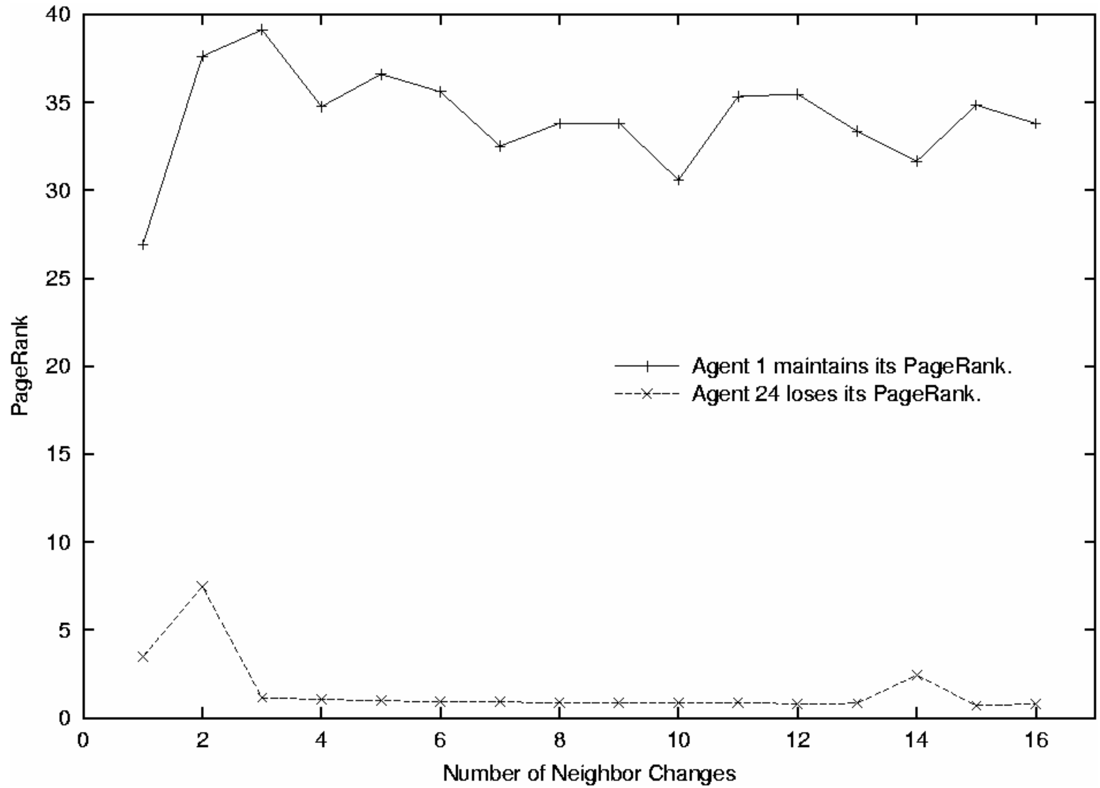
Providers:	Choose the best *m* agents whose expertise matches the agent's interests
Sociables:	Choose the most sociable *m* agents of its acquaintances
Weighted Average:	Choose the best *m* based on weighing both the expertise and the sociability of the acquaintances

Effect of Neighbor Selection Policies



Choosing sociables does not help authorities to emerge

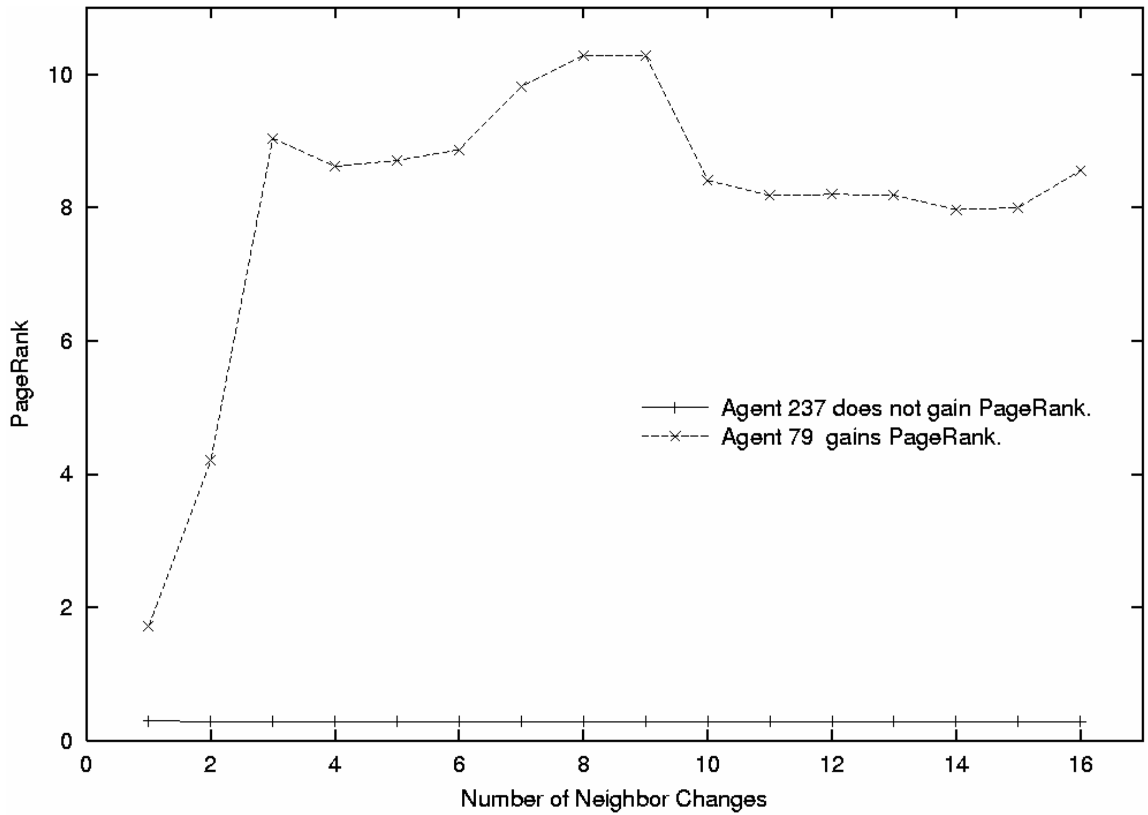
Decreasing Expertise; Preferring Sociables



Given: agents 1 and 24 lose their expertise

Evolution: yet, agent 1 remains authoritative because of its sociability

Increasing Expertise; Preferring Experts



Given: agents 79 and 237 become experts

Evolution: yet, agent 79 does not become authoritative because it is pointed to by only a few

Conjecture: After a population becomes stable,

- If agents prefer experts, then the winner need not take it all (i.e., a new expert can eventually become authoritative)
- If agents prefer sociables, then the winner takes it all (i.e., a new expert does not become authoritative)

Literature

- Referral systems:
  - MINDS
  - ReferralWeb
- Service location
  - Directory services (WHOIS++, LDAP)
    - No modelling of other servers
    - Rigid referrals (if any)
  - Chord, CAN, Pastry:
    - Routing based on a distributed hash table
    - No support for autonomous or heterogeneous peers

Directions

- Practical
  - Reimplement MARS
  - Incorporate QoS
- Research
  - Domain ontologies
  - Policies
  - Virtual Organizations

MS Themes

PhD Themes

Ontologies

- An ontology is a knowledge representation of some domain of interest
  - Successful communication (or interoperation) presupposes agreement of ontologies
  - Currently: develop standard ontologies for each domain
    - Time consuming; fragile
    - Doesn’t scale; omits opinions

IEEE SUO;
Cyc;

Language-based approaches:
WordNet;
LDOCE

Consensus

- Referral systems are a decentralized way to achieve (or approximate) consensus
  - About services, as above
  - Potentially also about ontologies
  - Use social network to determine who is an authority in what topic
  - Find a way to combine their ontologies for those topics

Big challenge: how to convincingly evaluate the contribution

Logic-Based Policies

- Referral systems appear to work, but how can
  - We be sure nothing bad will happen
  - An administrator or user configure such systems
- Use declarative policies to capture the agents' behavior
  - Use logic programming to develop the agents

Early stages: Udupi & Singh

Virtual Organizations

- Organizations of autonomous, heterogeneous parties collaborating some computational task
  - Common in scientific computing
  - Emerging in business settings
- Challenges VOs face
  - Interoperation of information resources as in other systems
  - Governance regarding allocating resources

Challenge to combine commitments with referral systems

Key Ideas

- Decentralization is desirable
  - Leave the user in control
  - Provide bookkeeping support
- Reputation *in action*
  - Not separated from usage
  - Context provides meaning to pointers
- Interesting properties of clustering and emergence
- Intuitive model underlying link analysis

Backup Slides

AutTitle

- Text

Sidebar

Basic Experimental Setup

- Interests used to generate queries
- Query, answer, interest, and expertise are vectors from Vector Space Model where each dimension corresponds to a domain
- Dimension of the vectors is 4
- Sociability is scalar
- 400 agents, with 10 to 25% service providers
- 8 neighbors per consumer
- Initial neighbors picked randomly
- Reselect neighbors after every 2 queries
- 4 to 20 neighbor changes

Metrics

- Qualifications:
  - **Similarity**: A symmetric relation to measure how similar two vectors are

$$I_A \oplus I_B = \frac{e^{-\|I_A - I_B\|^2} - e^{-n}}{1 - e^{-n}}$$

  - **Capability**: An asymmetric relation to measure how much better a vector is compared to the other

$$Q \otimes E = \frac{\sum_{t=1}^{n} (q_t e_t)}{\sqrt{n \sum_{t=1}^{n} {q_t}^2}}$$

Metrics

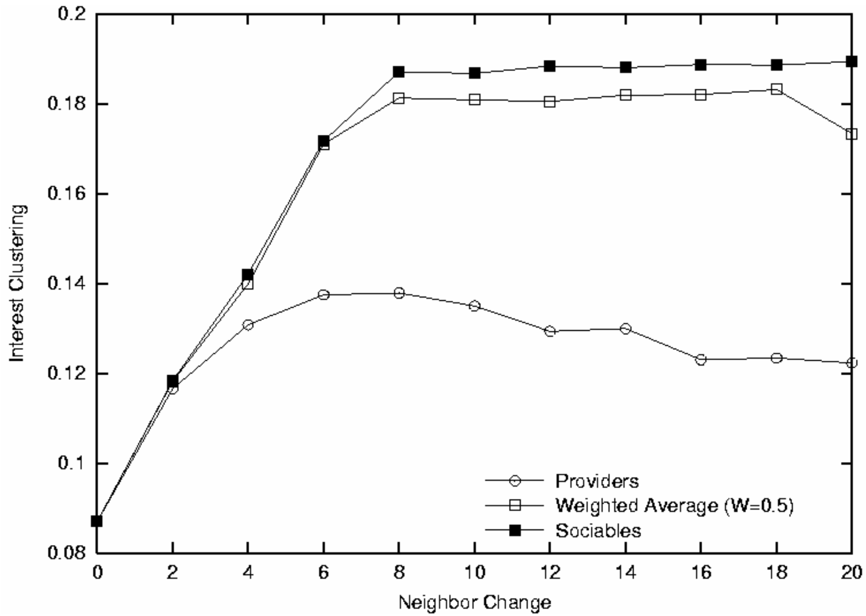
- Quality:
  - **Direct**: How close a match are the neighbors of an agent to it?

$$\frac{I_A \otimes E_B}{path(A,B)}$$

  - ***N*th Best**: Sort them and take the highest *n*th value. Each agent is represented by its *n*th best matching neighbor
- PageRank:

$$P(i) = d \sum_{j \in K_i} \frac{P(j)}{N_j} + (1 - d)$$

Clustering



Agents with similar interests

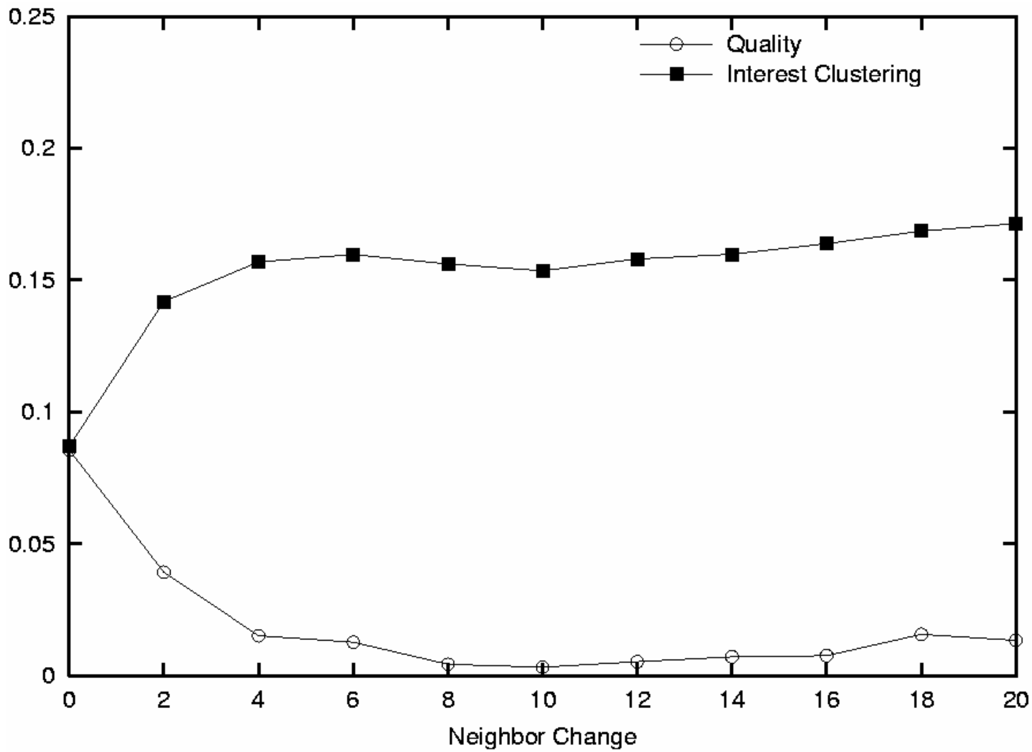
- May be looking for similar providers
- May give useful referrals
- Thus, will be considered sociable, and kept as neighbors

Measures how similar the neighbors of an agent are as well as how similar the agent is to its neighbors

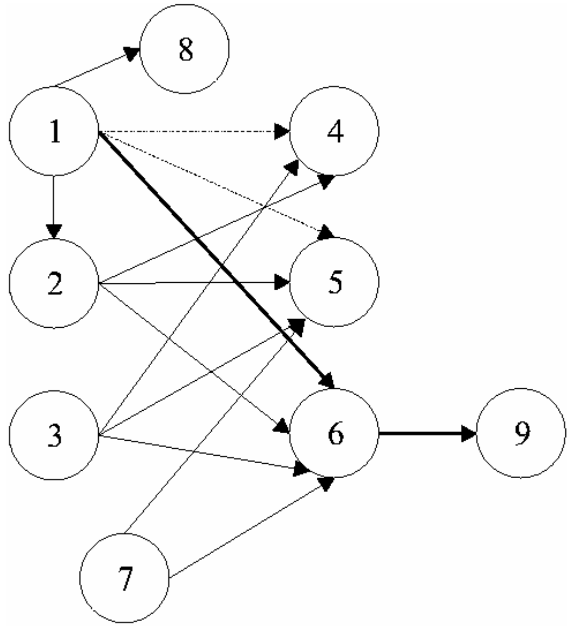
Sociability increases interest clustering

Clustering (2)

Result: Quality *decreases* when interest clustering increases



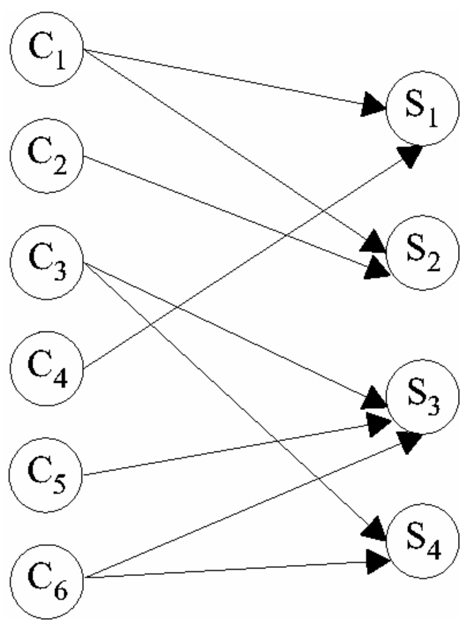
Co-Citation versus Referral Communities



- After running HITS, agents 4 and 5 are found to be authorities.
- Agent 1 generates a query.
- For bipartite communities, asks agents 4 and 5.
- For the referral communities, it poses the query to its choice of neighbors; in this case to agent 6 who gives a referral to agent 9.

Bipartite Communities
Referral Communities

Result: In a population where each agent exercises the *Providers* policy, if there are more providers than the number of neighbors an agent can have, then the graph converges into a bipartite graph



Bipartite Graphs

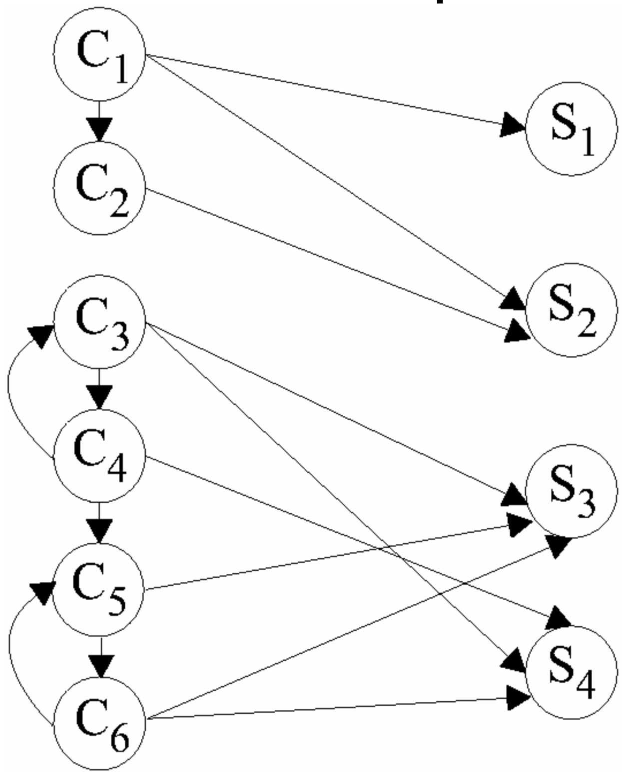
Weakly-connected components

Approximate how close a graph is to being bipartite:

Removing k edges

Removing k vertices

Result: In a population where each agent exercises the *Sociables* policy, the graph ends up with a number of weakly-connected components



Bipartite Graphs
Weakly-connected components

If there is more than one weakly-connected component, then there is at least one customer who will not be able to find a service provider

In-Degree Distributions

- Referral Policies
- Neighbor Selection Policies

Power Laws

- Rank each node from 1 to N, based on its in-degree
- The node n_i, with the highest in-degree $(I(n_i))$ will have $R(n_i) = 1$ and the node with the smallest in-degree (n_j) will have $R(n_j) = N$
- The distribution follows a power law if (for $\alpha < 0$)
  $R(x) = kI(x)^\alpha$
  $\log R(x) = \log k + \alpha \log I(x)$
- Zipf's Law is a power law where the slope is -1 and holds for
  - Number of people in a city (USA, India)
  - Occurrence of words in the English language

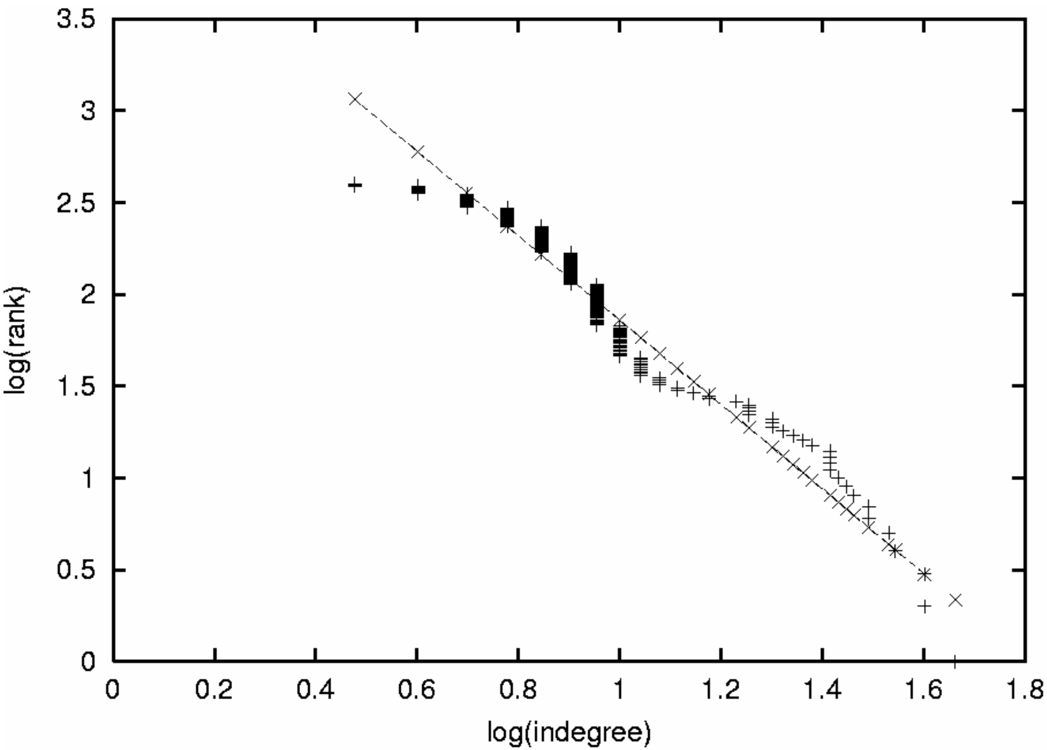
On Power-Law Relationships of the Internet Topology

M. Faloutsos
P. Faloutsos
C. Faloutsos
(SIGCOMM 1999)

Interacting Individuals Leading to Zipf's Law

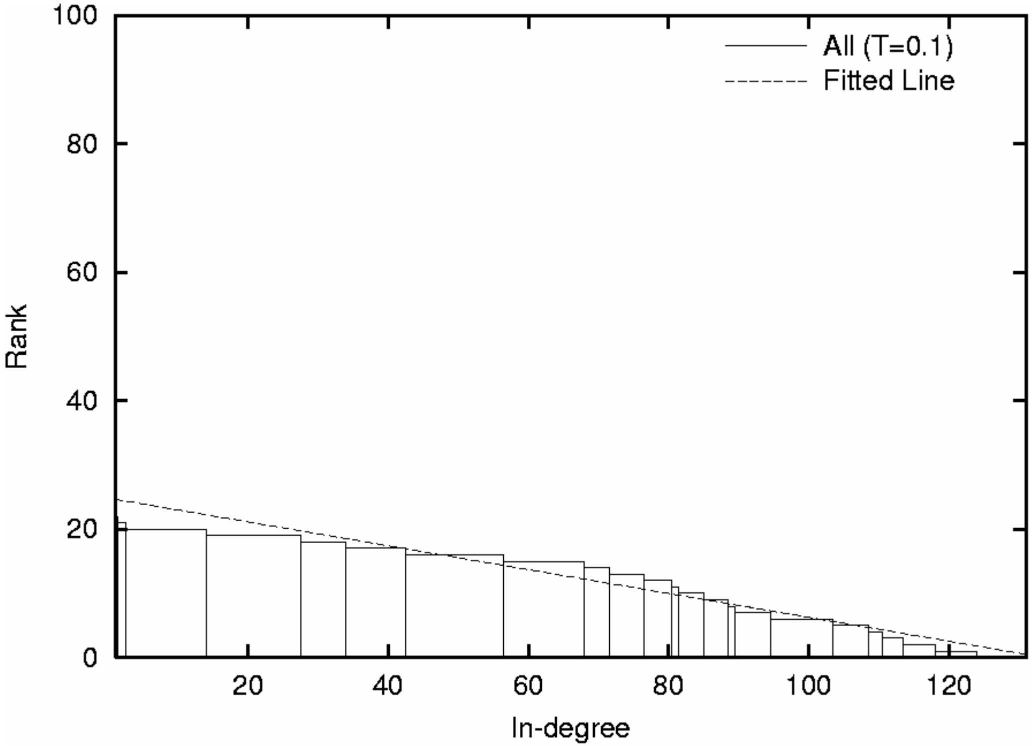
M. Marsili
Y. Zhang
(Physical Review Letters, 80(12), 1998)

Power-Law Distribution of In-Degree



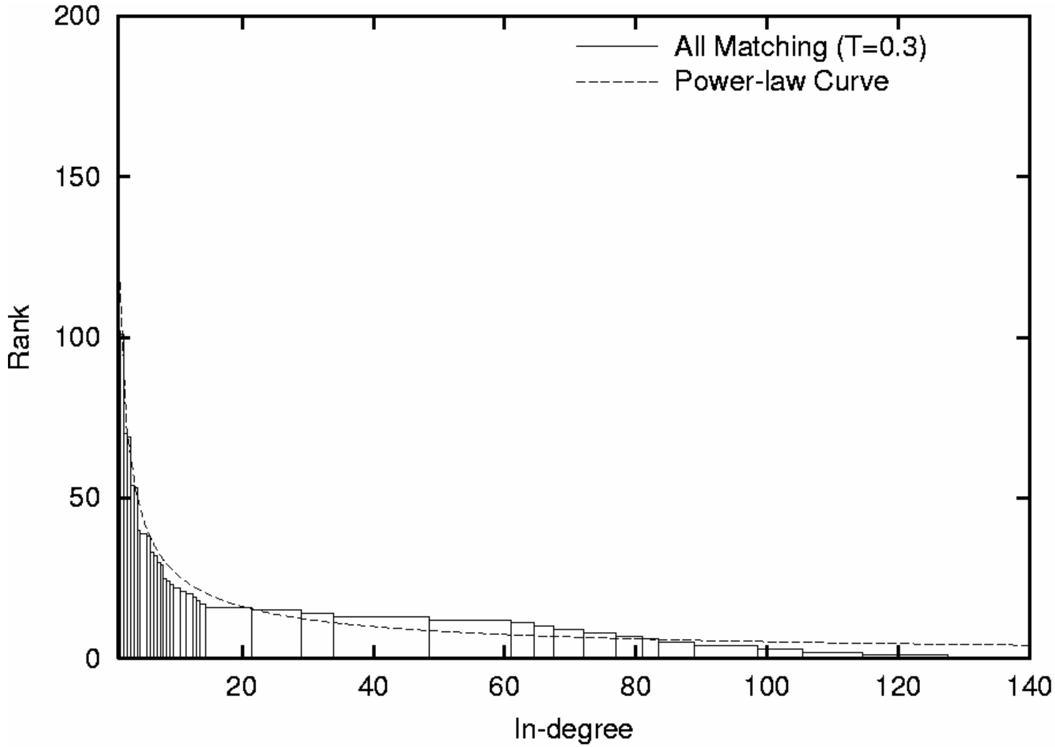
When agents are ranked based on their in-degree, the agent with the highest rank has a lot higher in-degree than the agent with the second rank, and so on

Agents Prefer Providers (1)



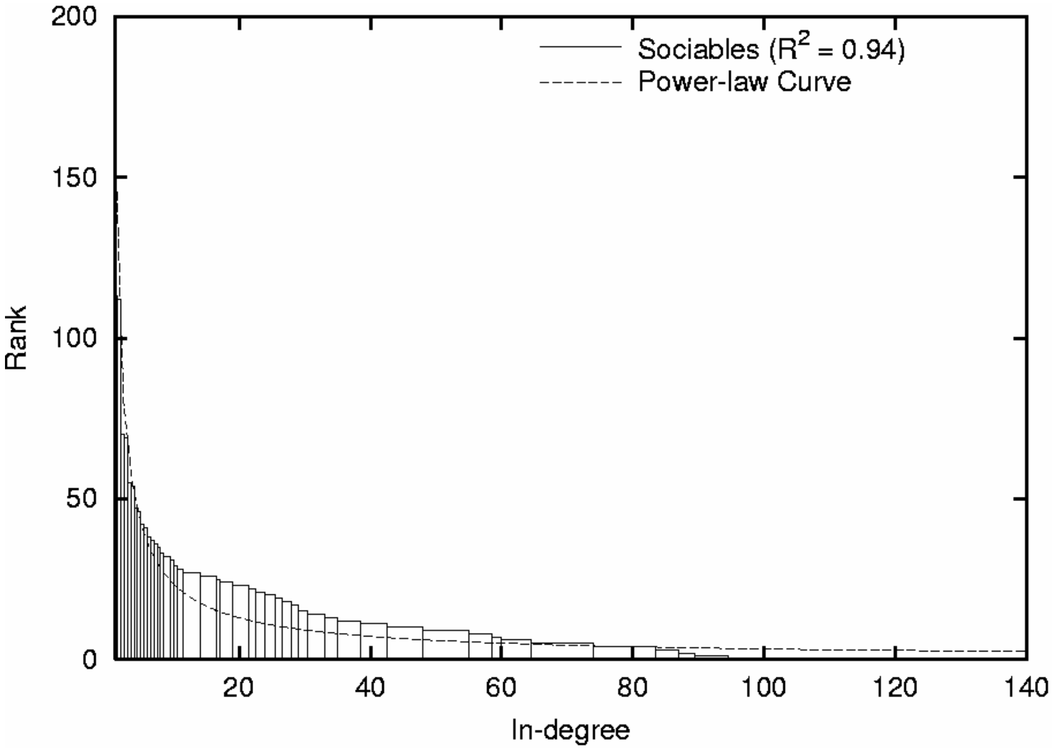
With nonselective referrals, when agents prefer providers, the in-degrees are shared among service providers

Agents Prefer Sociables (1)



1. With selective referrals, agents become locally sociable
2. In-degree distribution becomes a power-law

Agents Prefer Sociables (2)



Decreasing the selectivity of referrals decreases the fitness of the power-law

Discussion

- Reputation? What reputation?
  - Clearly being used in referrals
  - Clearly being built up or torn down
  - But not being computed as such (except for an after-the-fact study)
- Directions
  - Richer representations: transfer reputation across services
  - Protection against attacks: deception, collusion
- Implementation

Reputation

- Consider a society of principals, potentially each having opinions about the others
  - The opinions are applied implicitly in whether and how different parties do business with each other
- Someone's reputation is a *general opinion* about that party
  - Sometimes partially probed by asking others
  - Never explicitly fully aggregated, except in current computational approaches