Analysis of the Match Probabilities for the iTrust Information Network with Message Forwarding

Louise Moser and Michael Melliar-Smith

**University of California, Santa Barbara**

Introduction

- The iTrust system is a distributed and decentralized information publication, search and retrieval system, that is designed to defend against censorship of information in the Internet

- In this paper, we investigate the iTrust system with message forwarding, which spreads the responsibility of message distribution more widely across the nodes in the network and achieves greater scalability
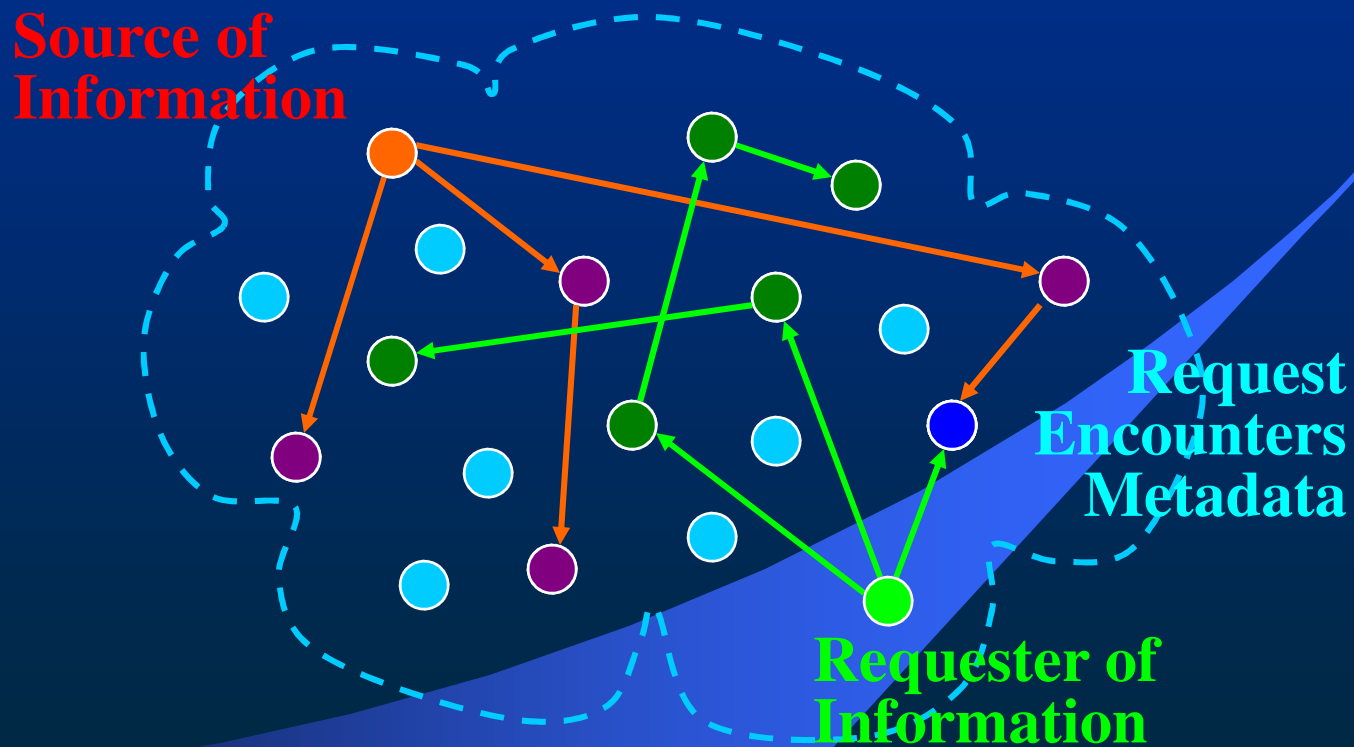
iTrust Information Network
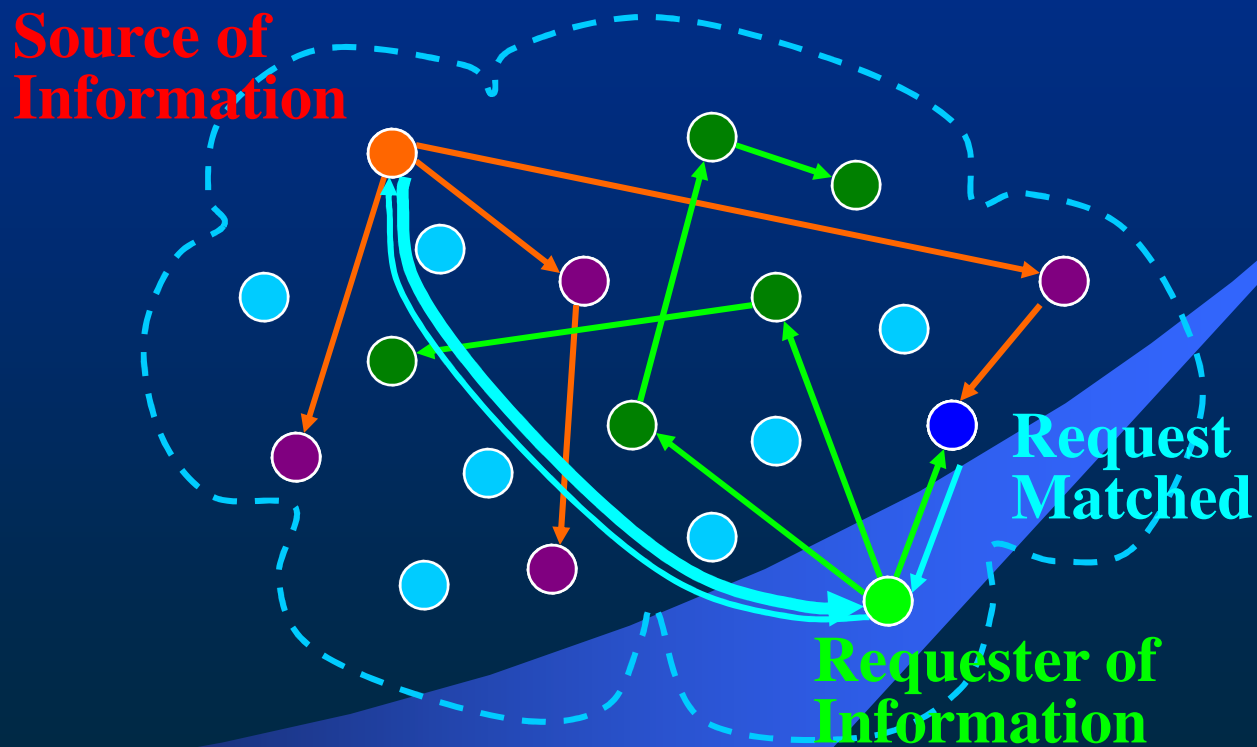## Distribution of Metadata
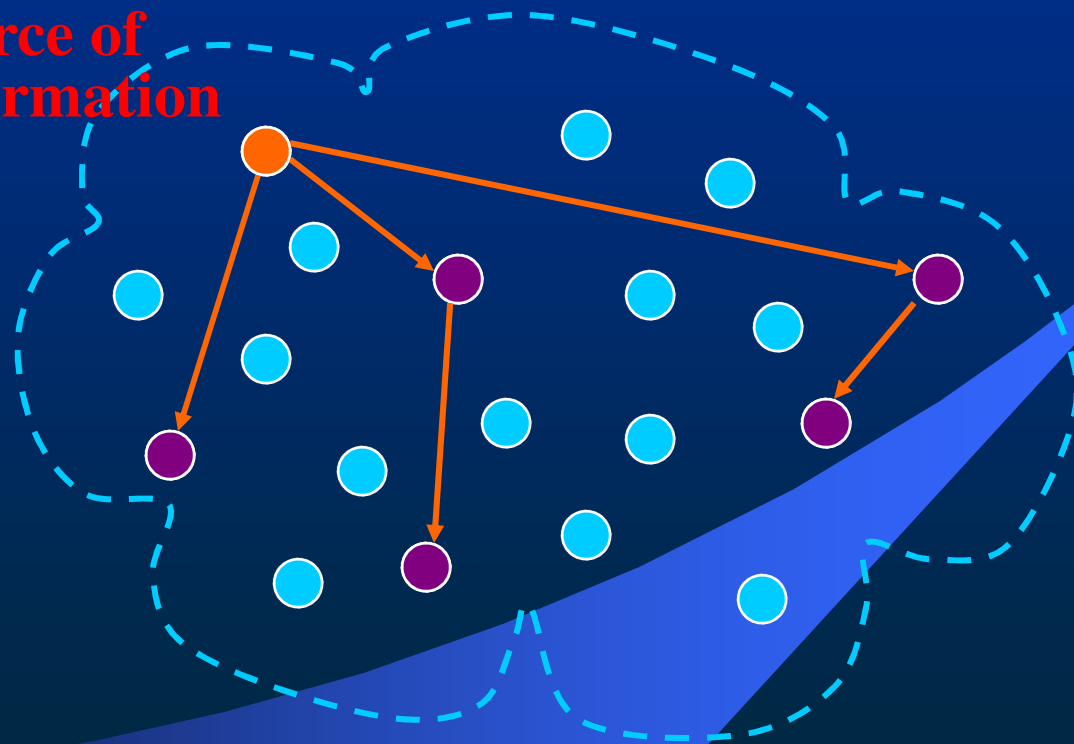
iTrust Information Network
Distribution of a Request

Source of Information

Request Encounters Metadata

Requester of Information

iTrust Information Network
Retrieval of Information

Source of Information

Request Matched

Requester of Information

iTrust Strategy

- Source node distributes its metadata to several nodes, chosen at random

- Requesting node distributes a request with keywords to several nodes, chosen at random

- Matching node sends URL to requesting node, which retrieves the information from the source node

- Parameters of iTrust
  - n: number of nodes in the network
  - m: number of nodes to which the metadata are distributed
  - r: number of nodes to which the requests are distributed
  - x: proportion of nodes that are operational

iTrust Match Probabilities

- The probabilistic analysis of iTrust without message forwarding is based on the hypergeometric distribution
- The probability of one or more matches is given by:

$$P(k \geq 1) = 1 - \frac{\dfrac{n-mx}{r} \quad \dfrac{n-mx-1}{r-1} \quad \cdots \quad \dfrac{n-mx-r+1}{1}}{\dfrac{n}{r} \quad \dfrac{n-1}{r-1} \quad \cdots \quad \dfrac{n-r+1}{1}}$$
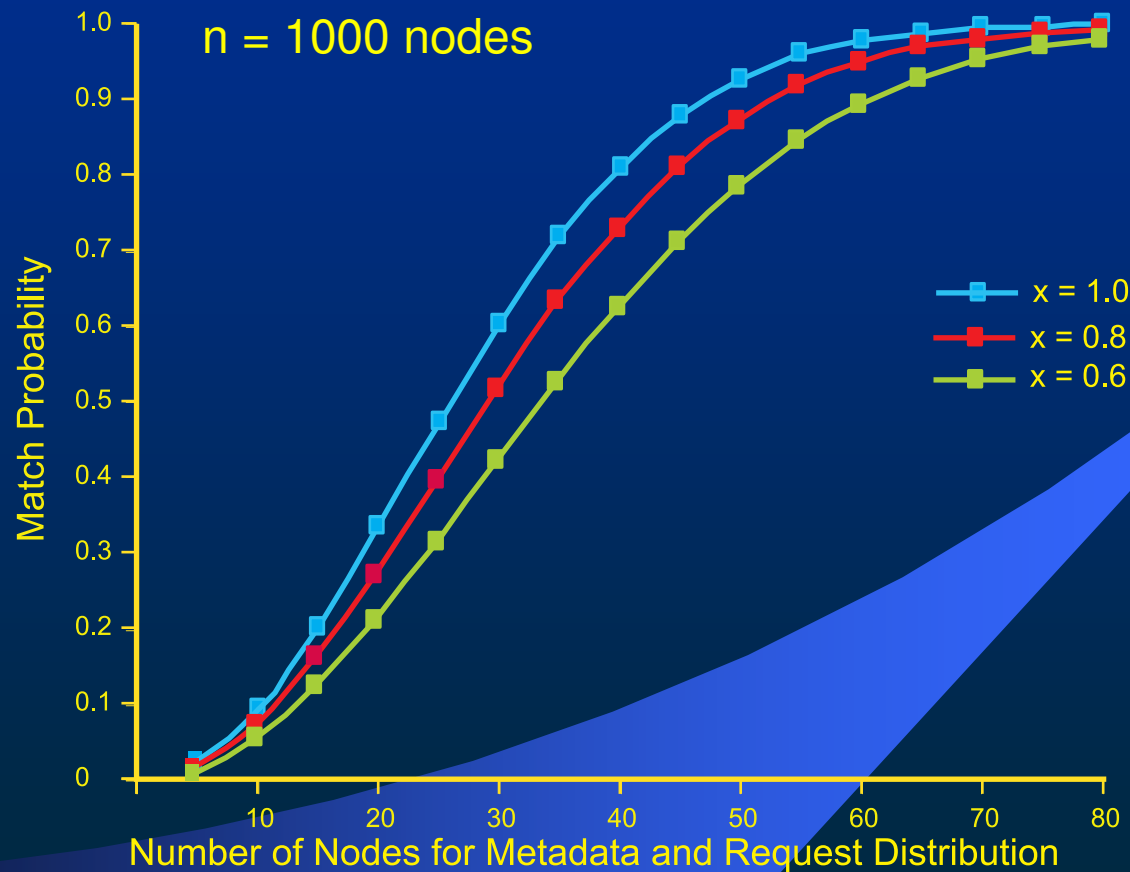
where mx + r ≤ n

iTrust Match Probabilities

- In an iTrust network containing n nodes,
  if the metadata and the requests are distributed to
  $2\sqrt{n}$ nodes then, as we have shown in [17],
  the probability of a match

  $$P(k \geq 1) \geq 0.9817$$

- In an iTrust network, we do not need to flood
  the entire network. Rather, we need to distribute
  the metadata and the requests to only $2\sqrt{n}$ nodes

iTrust Match Probabilities

n = 1000 nodes
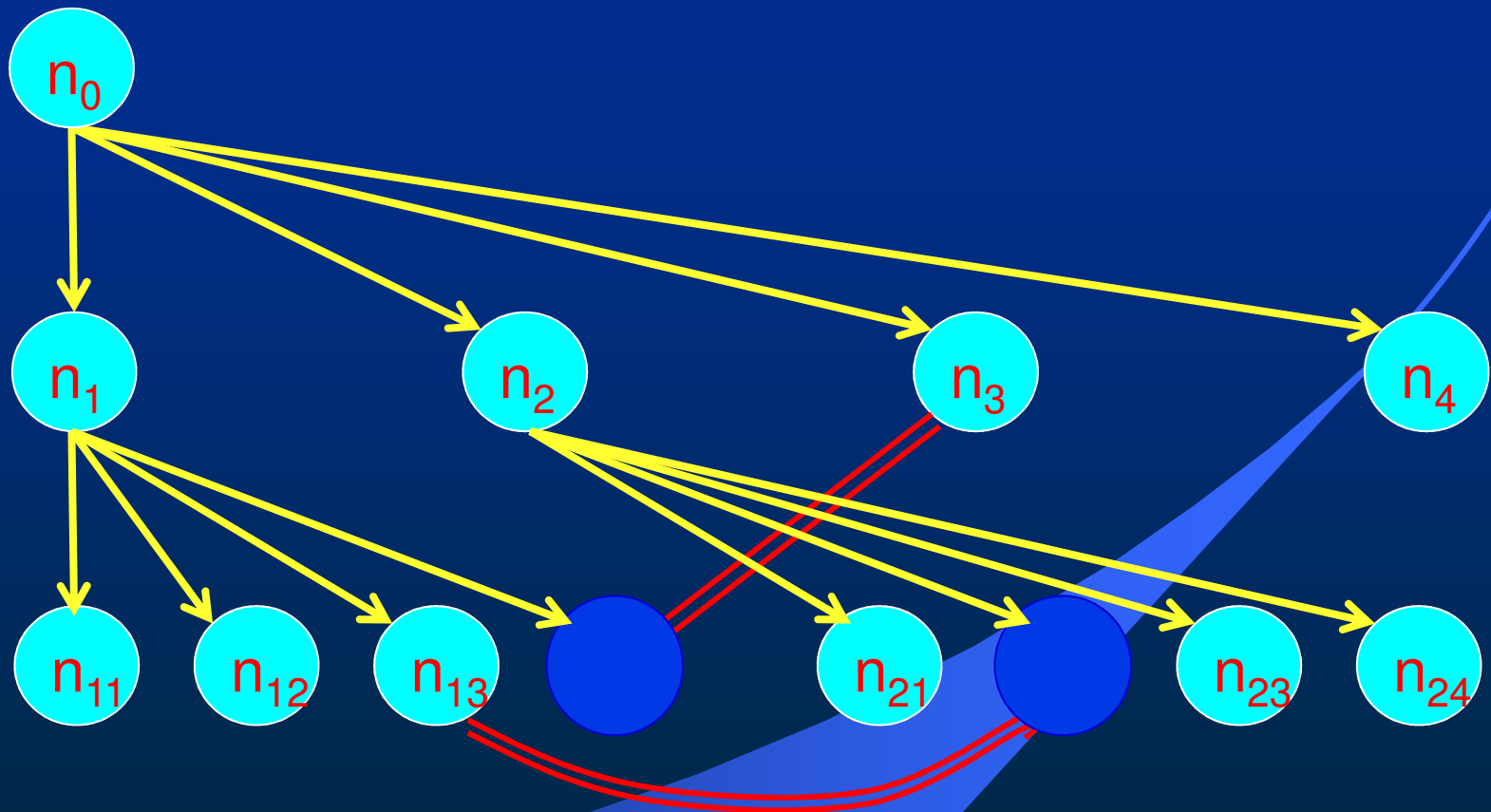
Match Probability (y-axis)

Number of Nodes for Metadata and Request Distribution (x-axis)

Legend:
- x = 1.0
- x = 0.8
- x = 0.6

iTrust with Message Forwarding

- Source node transmits its message to several nodes
  - Each such node retransmits the message to other nodes, chosen at random
  - Spreads the load across multiple nodes
- Parameters of iTrust with message forwarding
  - n: number of nodes in the network
  - c: number of nodes to which a node forwards a message
  - f: probability with which a node forwards a message
  - ℓ: number of levels of message forwarding

iTrust with Message Forwarding
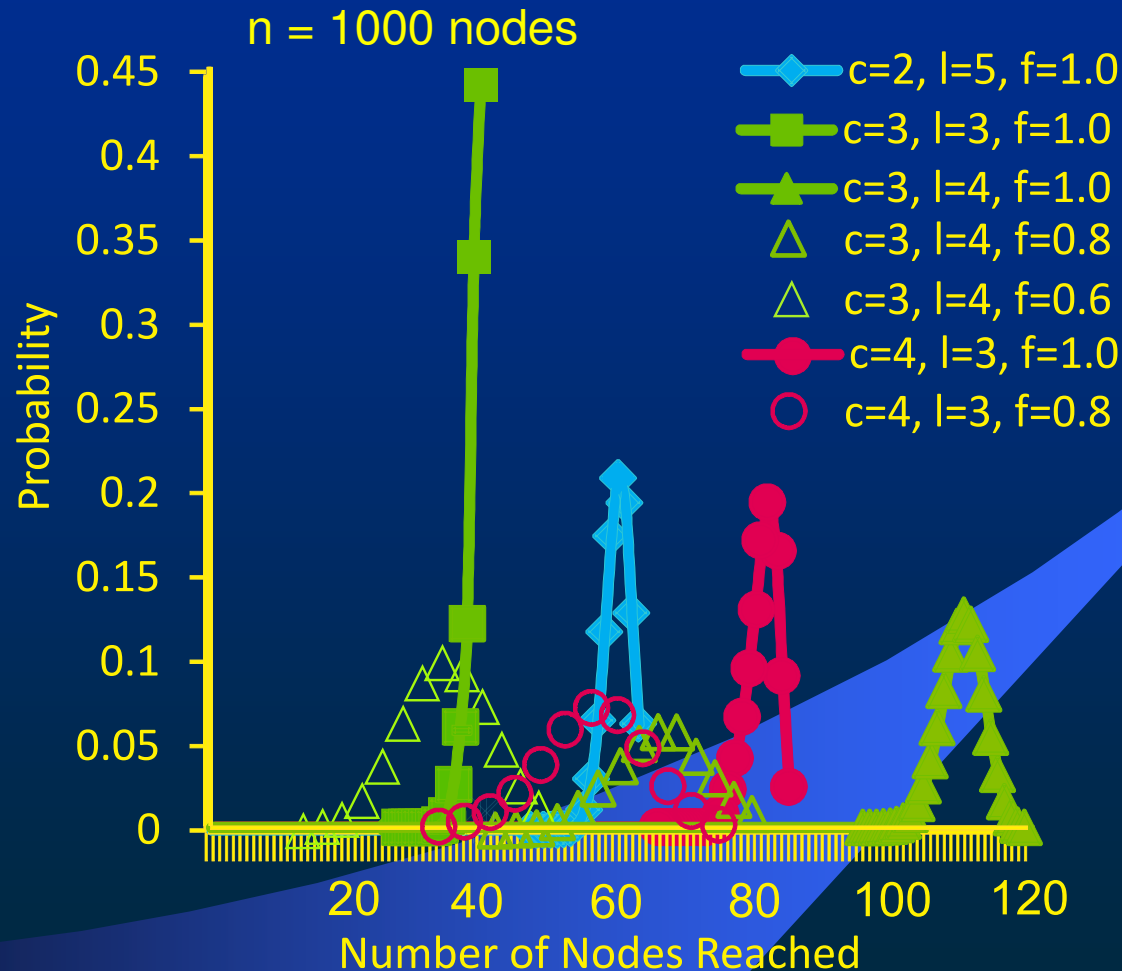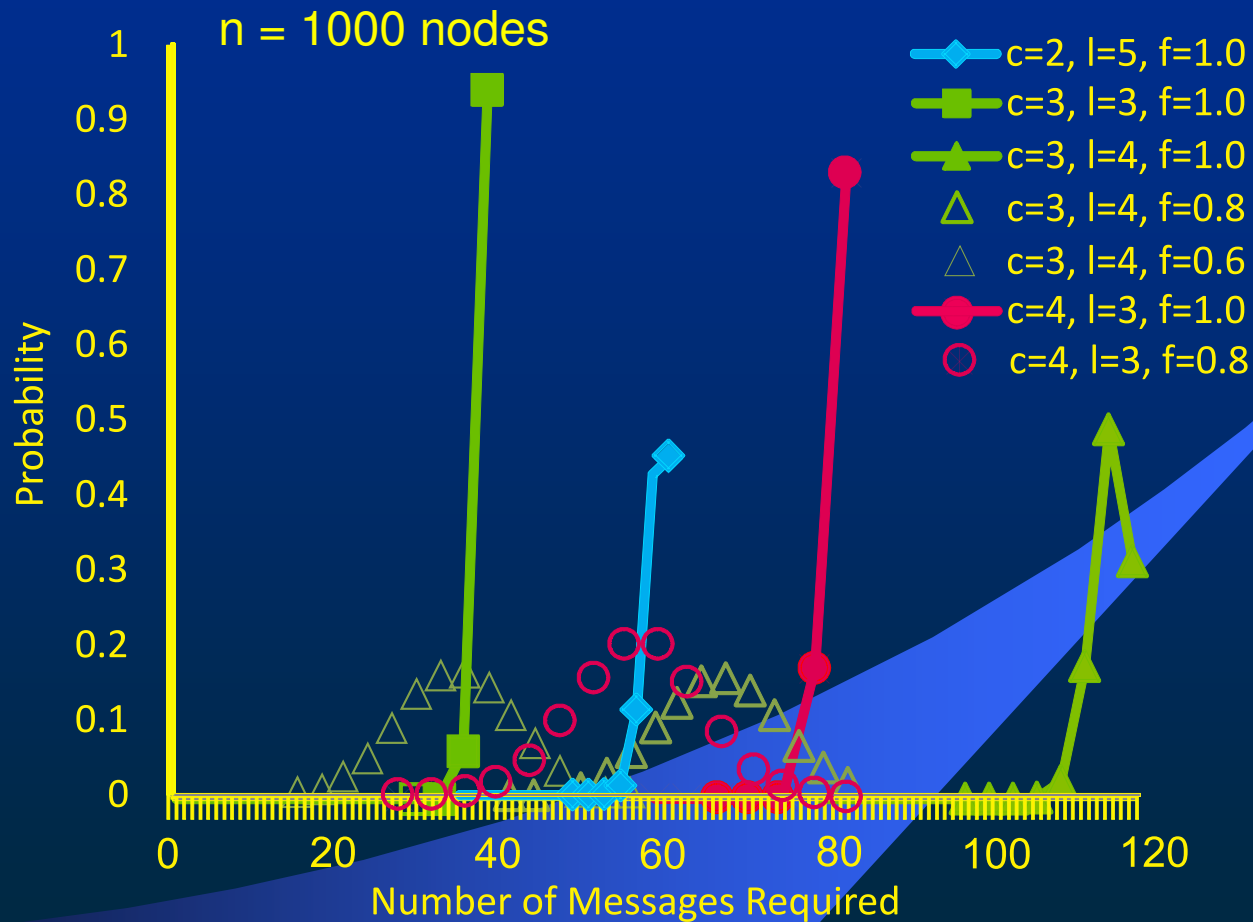
iTrust with Message Forwarding

- The probabilistic analysis of iTrust with message forwarding is based on:

  - probability of a match $P(k \geq 1)$

  - probability density functions pdf[m] and pdf[r], given by the algorithm that we presented in [18], for the number of nodes reached when forwarding the metadata to m nodes and the requests to r nodes

- In particular:

```
matchProb = 0.0
for m = c+1 to n do
    for r = c+1 to n do
        matchProb = matchProb + P(k ≥ 1) x pdf[m] x pdf[r]
```

pdfs for Nodes Reached
iTrust with Message Forwarding

n = 1000 nodes

Legend:
- c=2, l=5, f=1.0
- c=3, l=3, f=1.0
- c=3, l=4, f=1.0
- c=3, l=4, f=0.8
- c=3, l=4, f=0.6
- c=4, l=3, f=1.0
- c=4, l=3, f=0.8

Y-axis: Probability
X-axis: Number of Nodes Reached

pdfs for Messages Required
iTrust with Message Forwarding

n = 1000 nodes

Probability

Number of Messages Required

Legend:
- c=2, l=5, f=1.0
- c=3, l=3, f=1.0
- c=3, l=4, f=1.0
- c=3, l=4, f=0.8
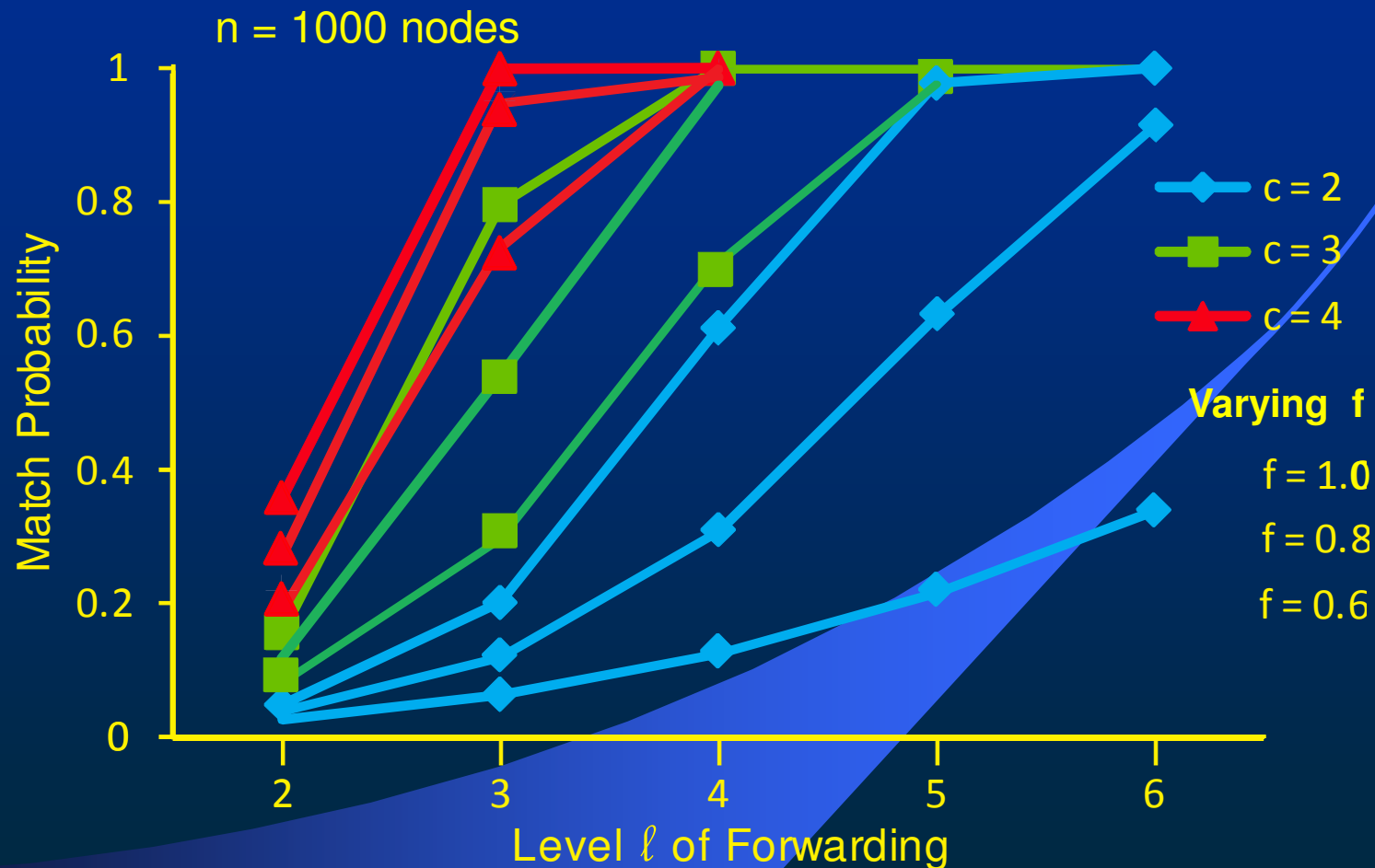- c=3, l=4, f=0.6
- c=4, l=3, f=1.0
- c=4, l=3, f=0.8

Match Probabilities
iTrust with Message Forwarding
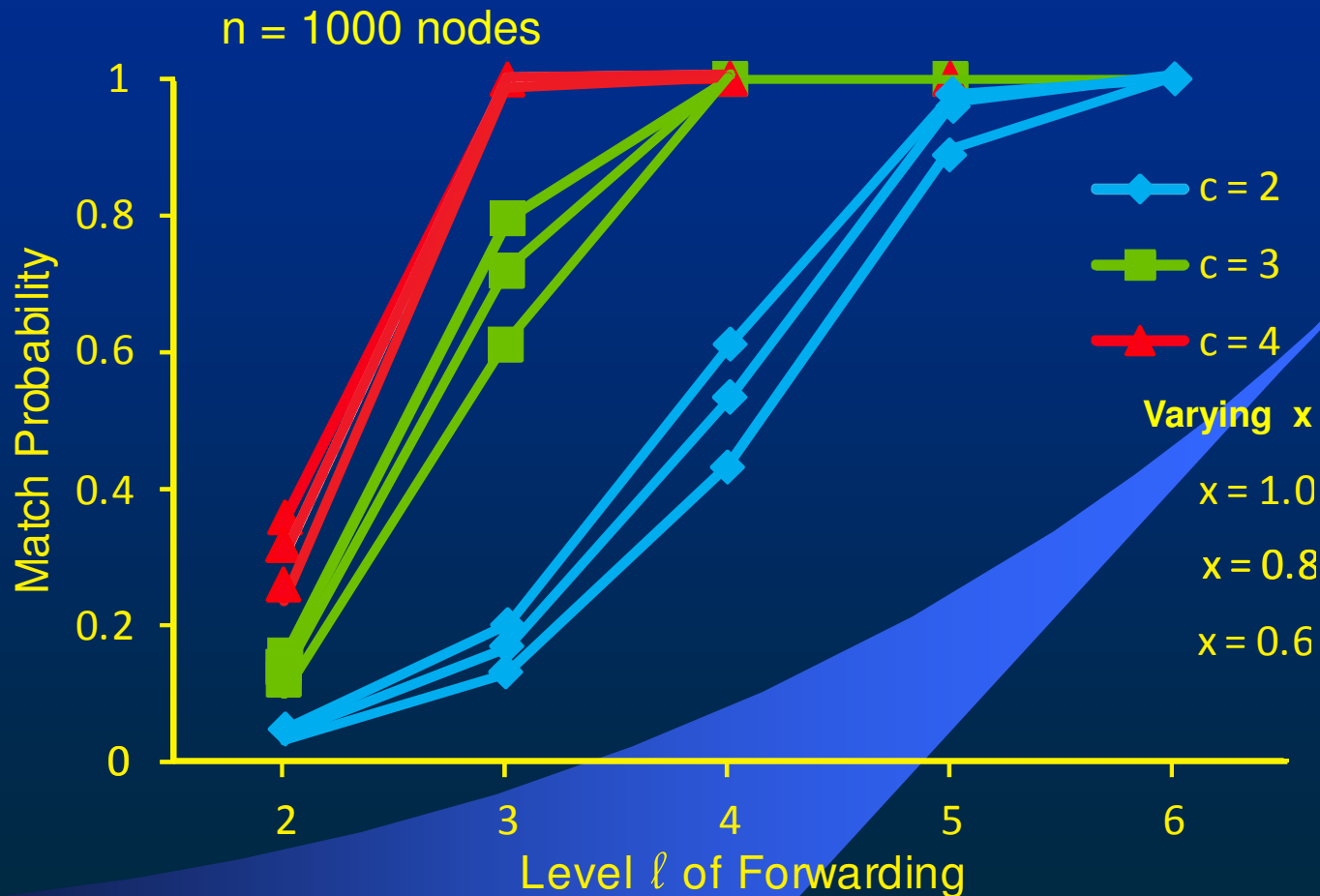
Match Probabilities
iTrust with Message Forwarding

Related Work

- S. M. Hedetniemi, S. T. Hedetniemi and A. L. Liestman, "A survey of gossiping and broadcasting in communications networks," *Networks*, vol. 18, pp. 319-349, 1988

- D. Shah, "Gossip algorithms," *Foundations and Trends in Networking*, vol. 3, no. 1, pp. 1-125, 2008

- Q. Lv, P. Cao, E. Cohen, K. Li and S. Shenker, "Search and replication in unstructured peer-to-peer networks,'' *Proc. IEEE Intl. Conf. Supercomputing,* June 2002, pp. 84-95

- R. A. Ferreira, M. K. Ramanathan, A. Awan, A. Grama and S. Jagannathan, "Search with probabilistic guarantees in unstructured peer-to-peer networks,'' *Proc. 5th IEEE Intl. Conf. Peer-to-Peer Computing*, Aug. 2005, pp. 165-172

Conclusions

- In iTrust, relatively small values of c and ℓ with f = 1.0 result in the metadata and requests being distributed to 2 $\sqrt{n}$ nodes and, thus, a high match probability and reasonable message cost

- With f < 1.0, the number of nodes reached and the match probabilities exhibit much greater variabilility with detrimental effects on the match probabilities

- Thus, in iTrust, it is preferable to adjust c and ℓ and to keep f = 1.0

Future Work

- Investigate the match probabilities of the iTrust information network with message forwarding, in networks that are not completely connected

- Many networks, particularly, social networks, contain neighborhoods that are completely connected locally but that have few connections to other neighborhoods

- Such networks can be very large, without each node knowing about a large number of other nodes

- Forwarding within, and between, neighborhoods can achieve scalability for the iTrust information network

Questions?   Comments?

- Louise Moser - moser@ece.ucsb.edu
- Michael Melliar-Smith - pmms@ece.ucsb.edu

This work was supported in part by
the US National Science Foundation
NSF CNS 10-16193