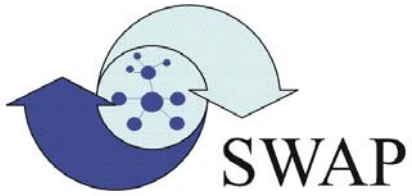


Topic Communities in P2P Networks

Steffen Staab

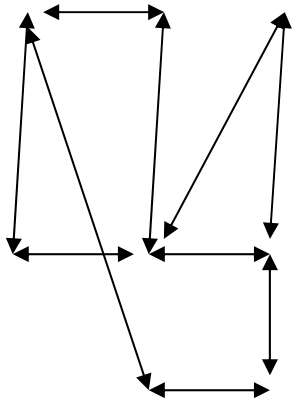
Joint work with A. Löser (IBM), C. Tempich (AIFB)

SNA@ESWC 2006

Budva, Montenegro, June 12, 2006

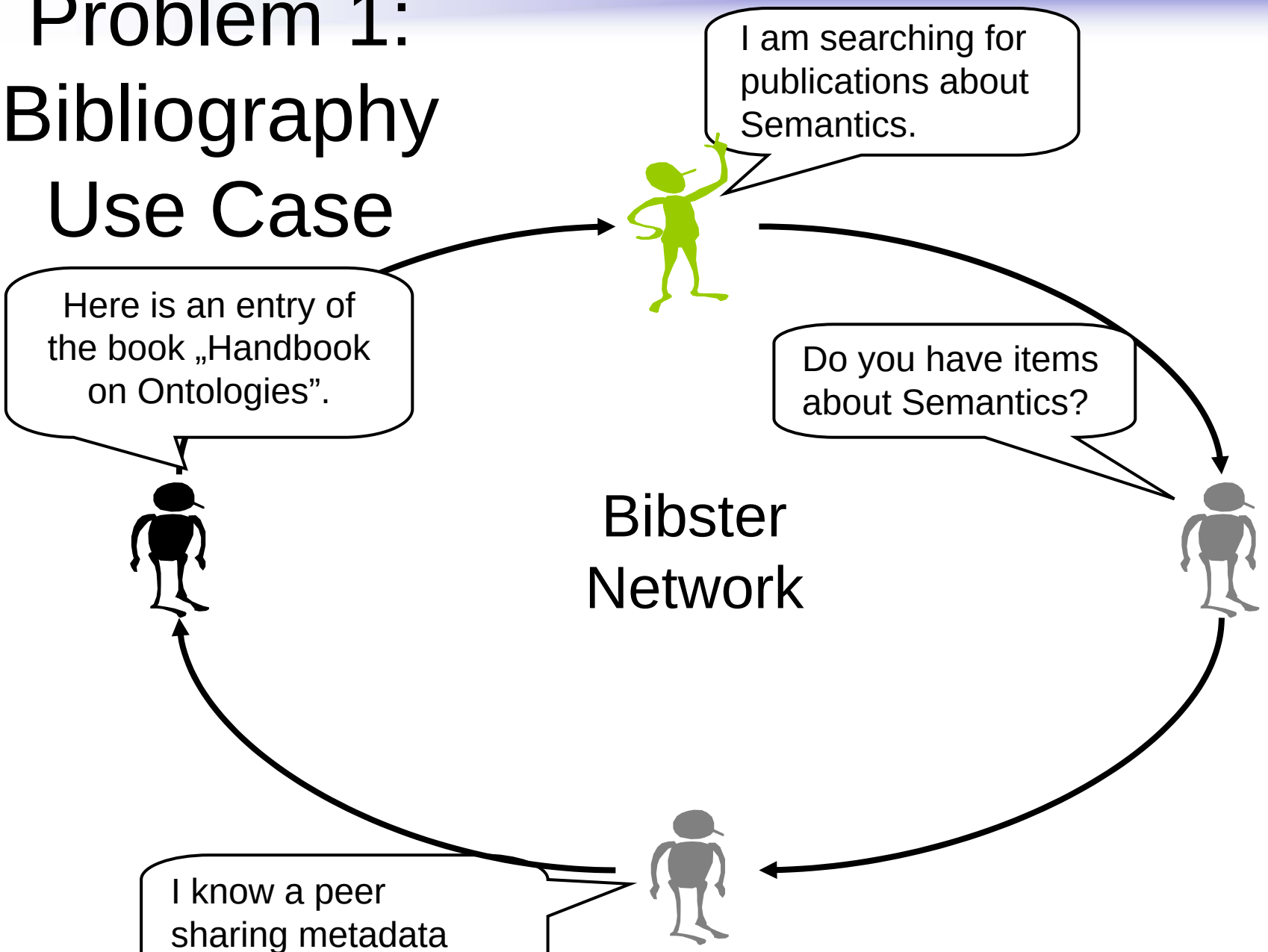
Two opposite challenges when considering Social Networks

- Analysis
  - Nodes/Agents
  - Links/Communication
  - Time
  - Topics
  - Causal Models,…

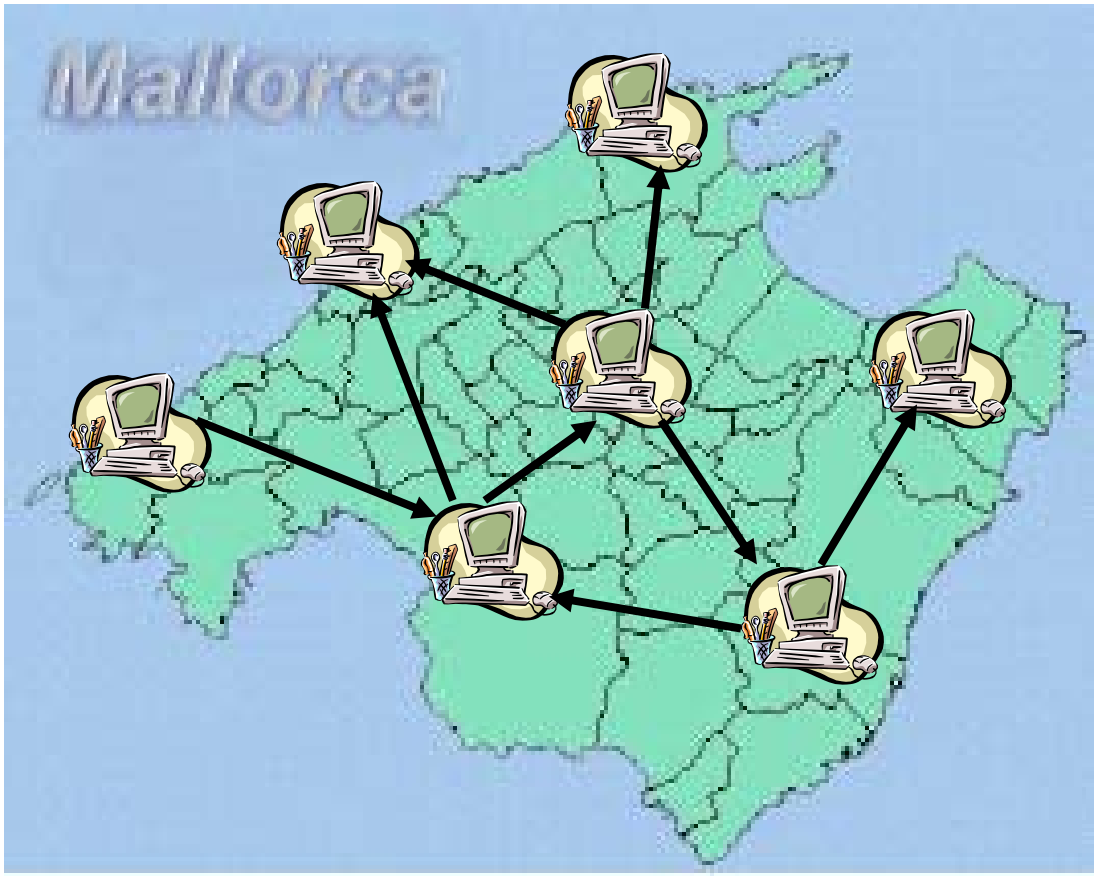


- Construction
  - Nodes/Agents
  - Construct Links
  - Live in time
  - Talk about topics
  - Influence others
  - …

Problem 1: Bibliography Use Case



Problem 2: Virtual Organization



- Tourism sector in the Balearic Islands shows:
  - Autonomy of individual actors, and
  - the need for cooperation of these actors.
- The ‘Balearic product’ can only succeed in the market if the actions of all the parties are well coordinated.
- We may talk of a Virtual Organisation in the BI created by shared interests and mutual dependencies.

Basic idea: Shortcut Creation based on Social Network Metaphors

"A question is asked to a person

Query Dependent

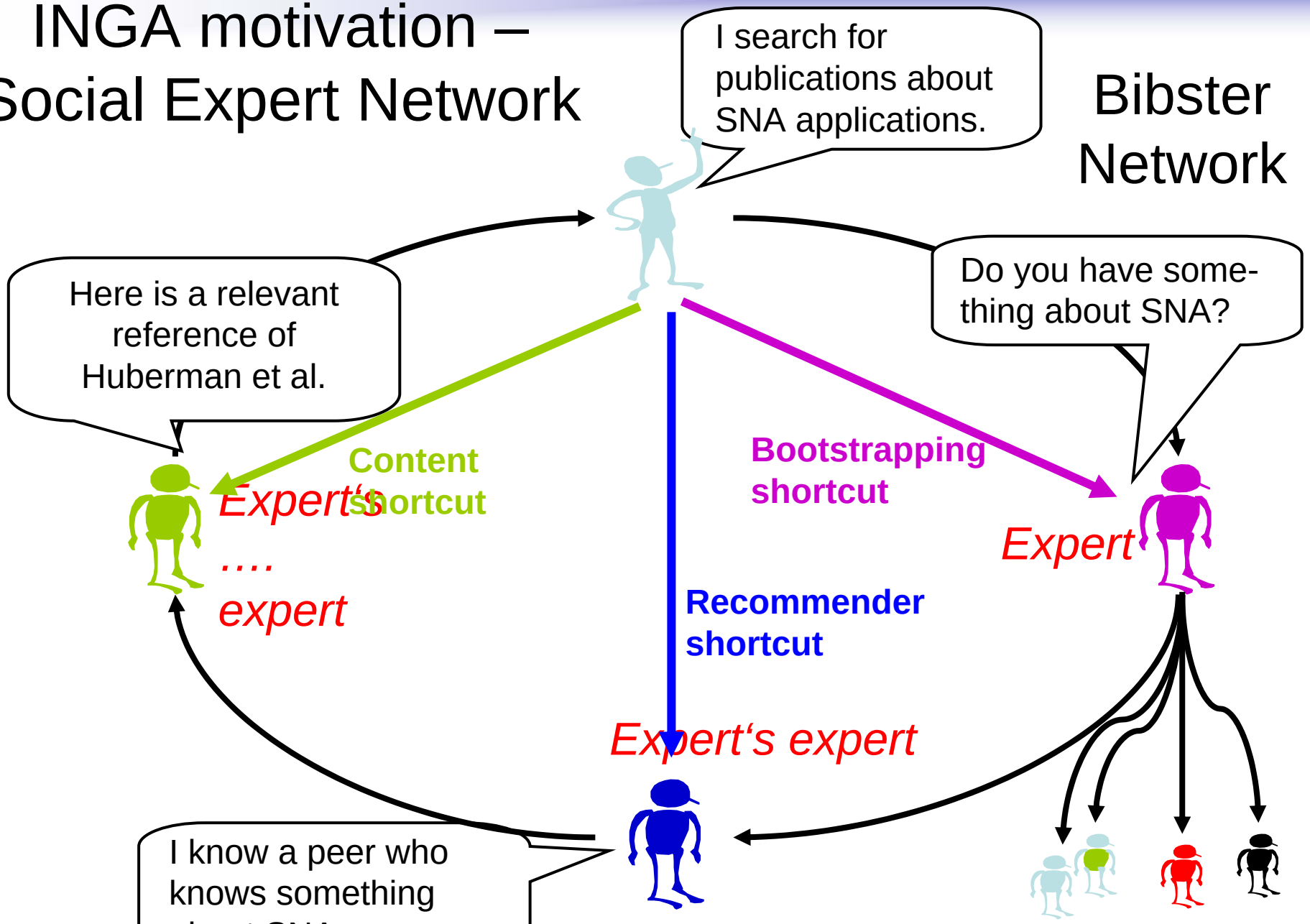
1. that has **answered** the question in the past successfully (*Content Provider).*

2. that has **asked** a similar question in the past successfully (*Recommender).*

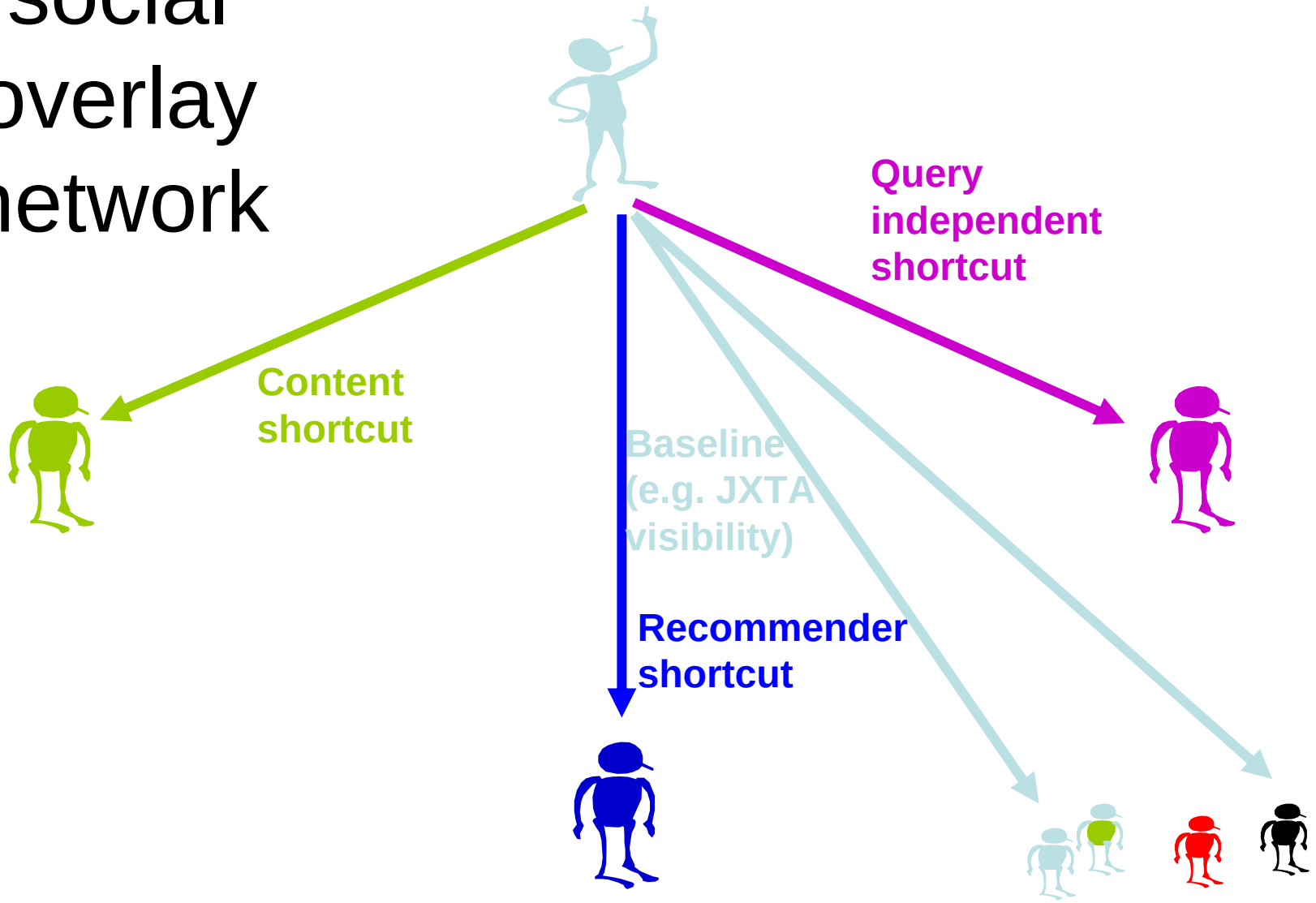
Query Independent

3. that has **established** a good **network** to other persons over several domains. Such persons form our *Bootstrapping Network.*

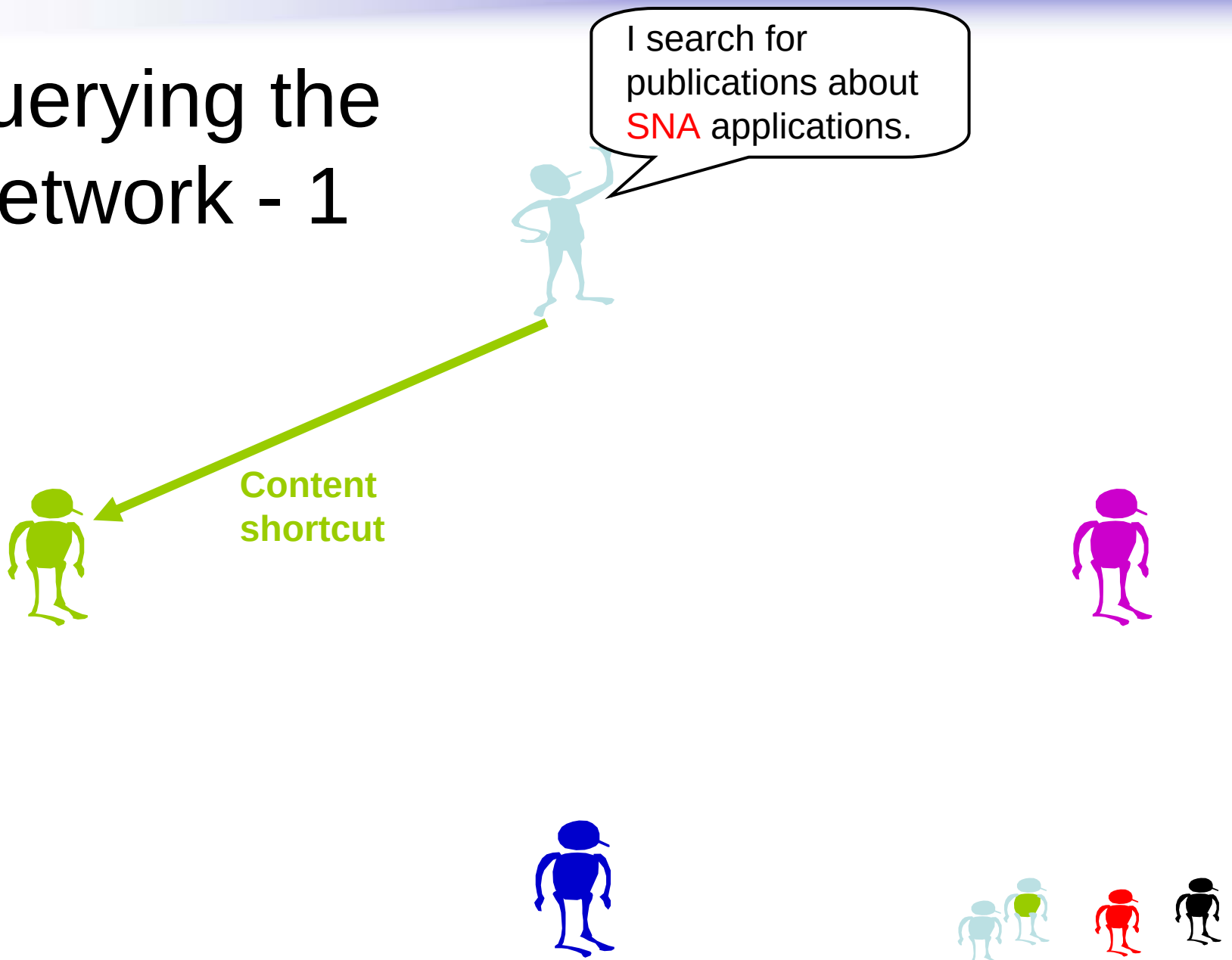
INGA motivation – Social Expert Network



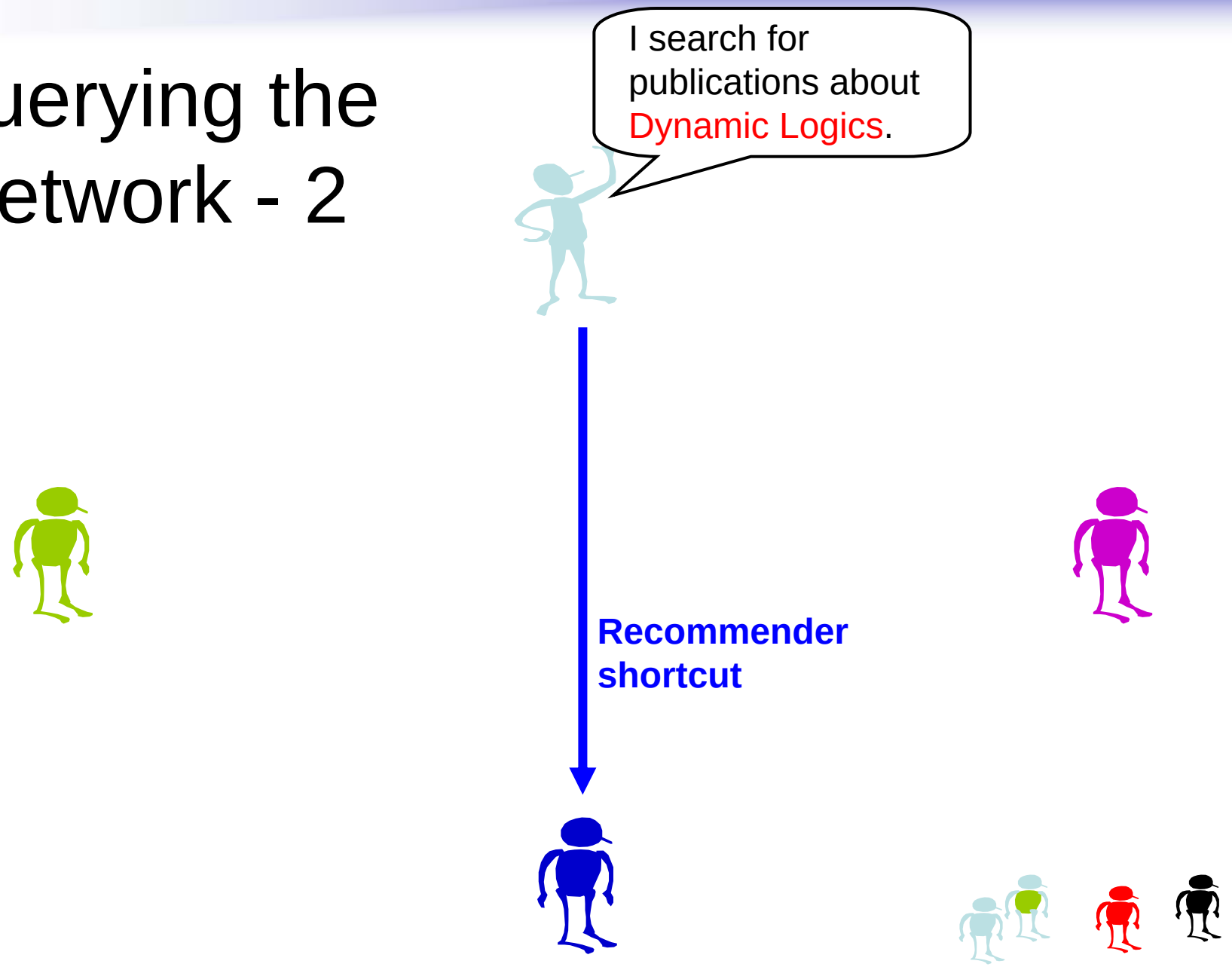
Semantic social overlay network



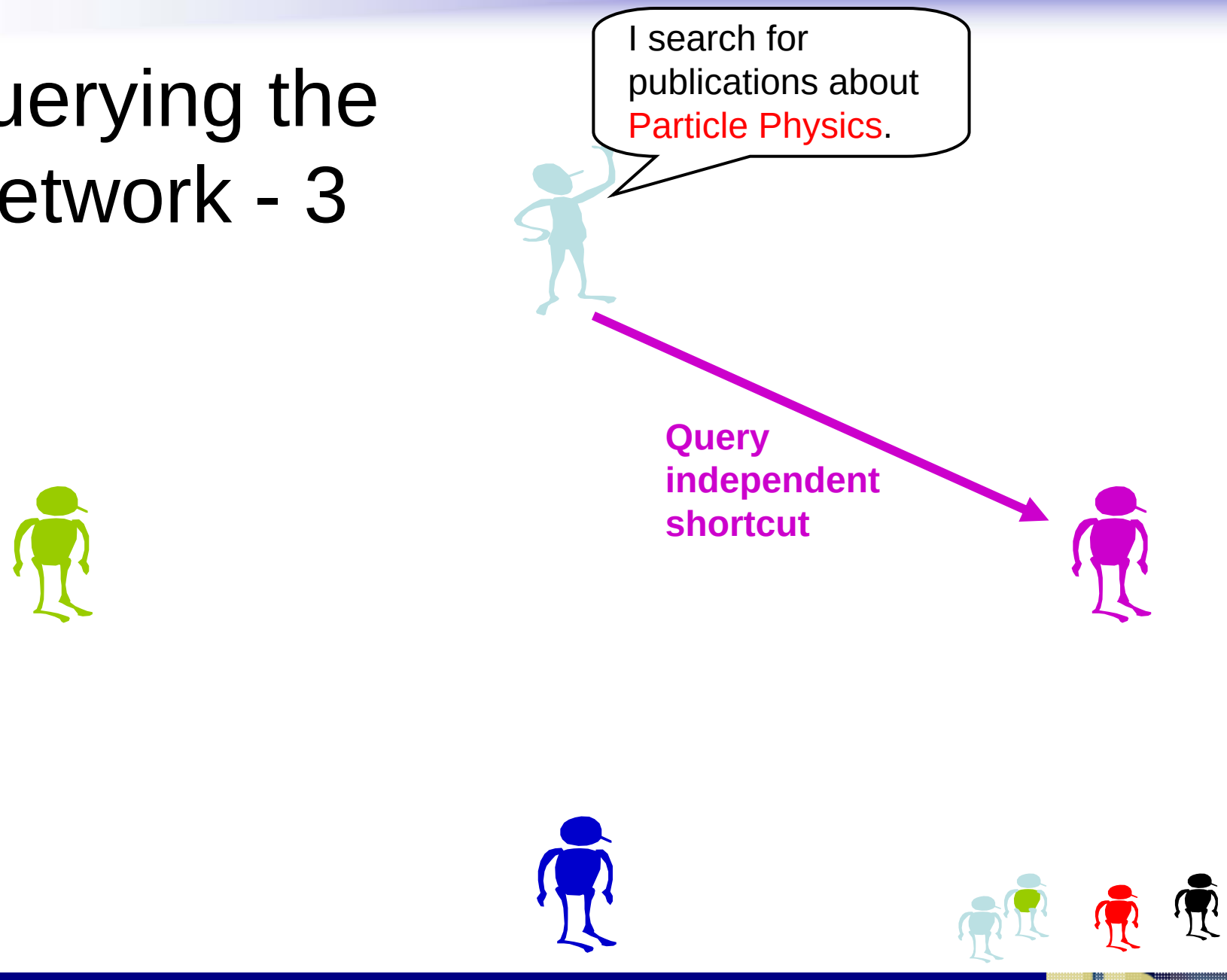
Querying the network - 1



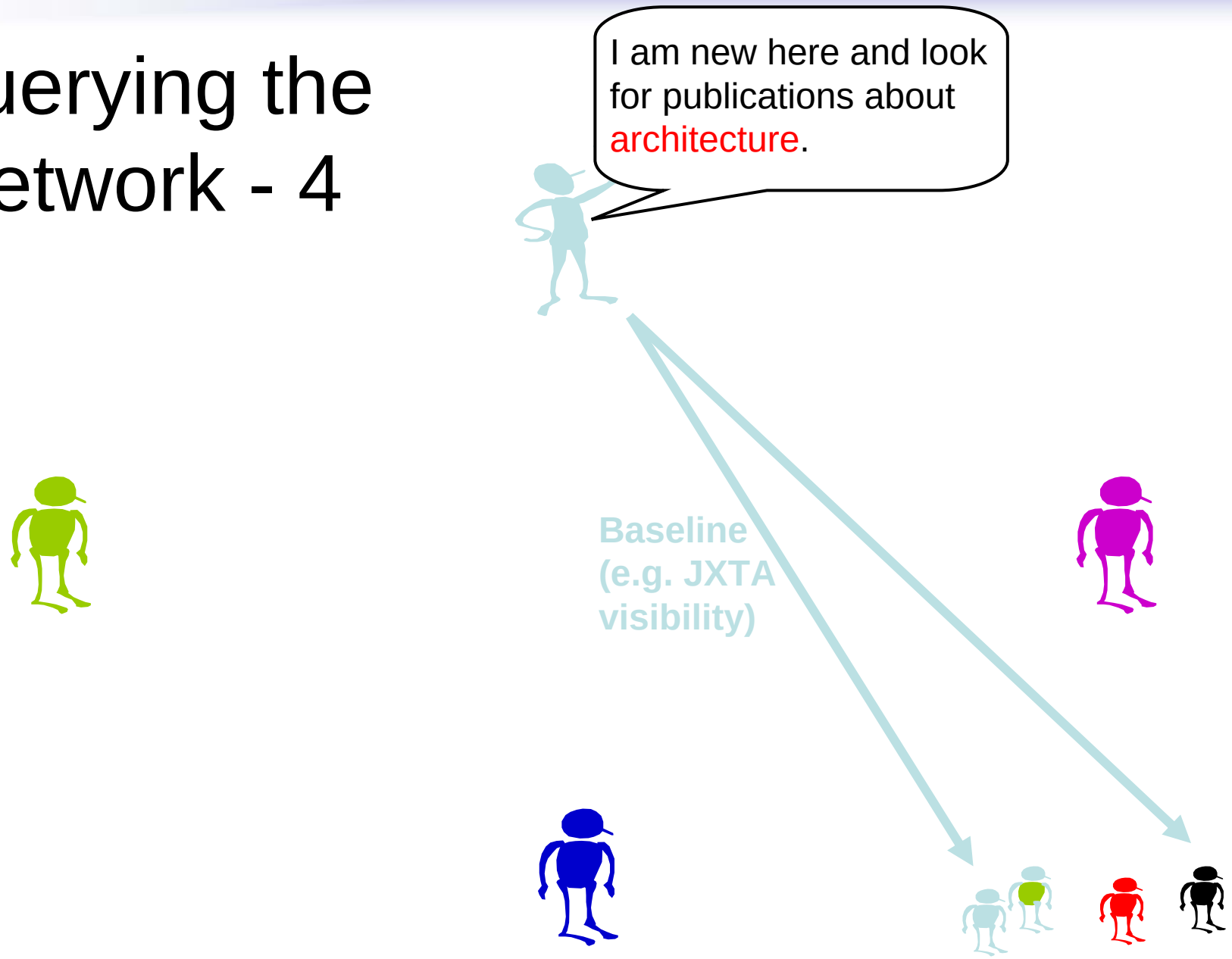
Querying the network - 2



Querying the network - 3



Querying the network - 4



Build content shortcut index

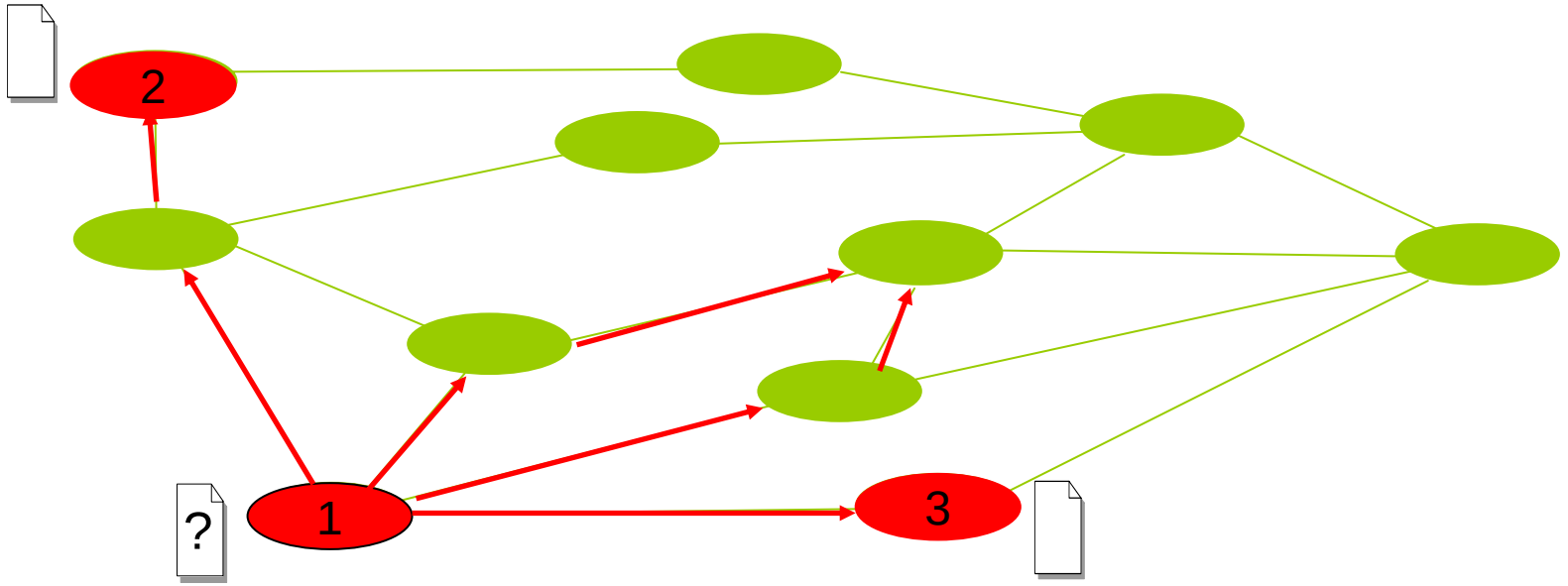
1. Send query using most promising available layer of semantic overlay topology

2. Evaluate result of query

3. Update shortcut index

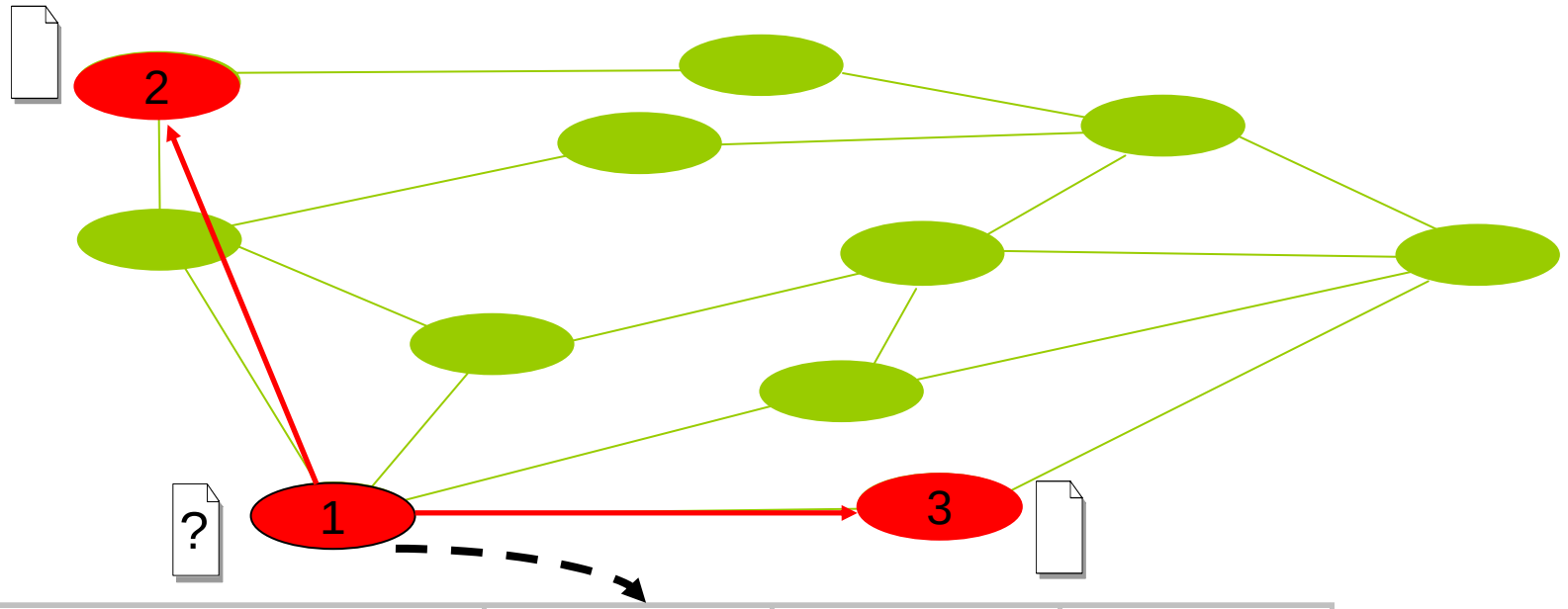
Use & Construct
Content Provider Shortcut Index

Query q1 from Peer p4: *Search materials about Semantic Web and Application of RDF*



Query:
/SemanticWeb/RDF

Use & Construct
Content Provider Shortcut Index



Query Q'	PID	Query Hits	SC Type
/SemanticWeb/RDF	2	100	C
/SemanticWeb/RDF	3	93	C

Build recommender shortcut index

Active

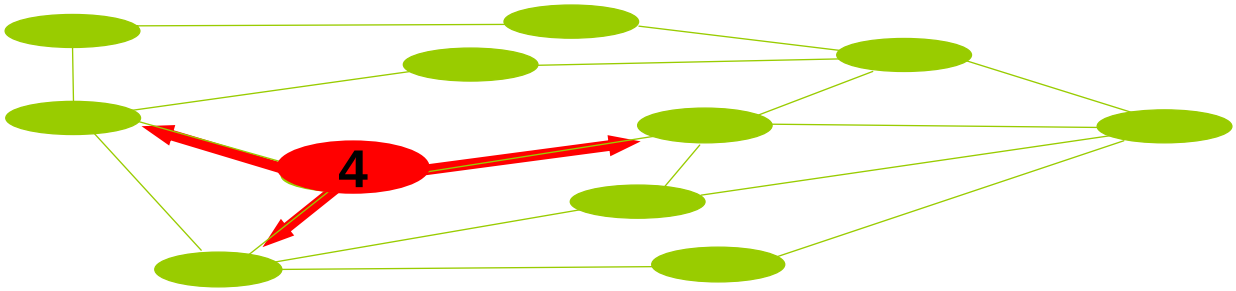
1. When answers are returned including the query message path:
   - The one butlast in the path is a recommender peer

Passive

1. Listen to incoming queries
2. Interest-based Indexing:
   If
   similarity(query,content$_i$) > threshold,
   then
   add P_q as Recommender Peer to Index

Create Recommender Shortcut Index (Active)

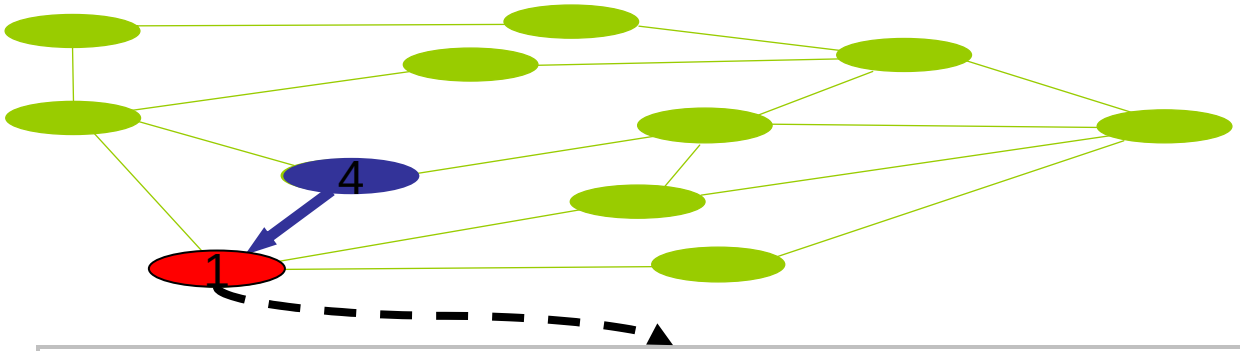
1. „Flood“ for existing Content Provider Short Cuts



- *Query /SemanticWeb/OWL*

Create Recommender Shortcut Index (Active)

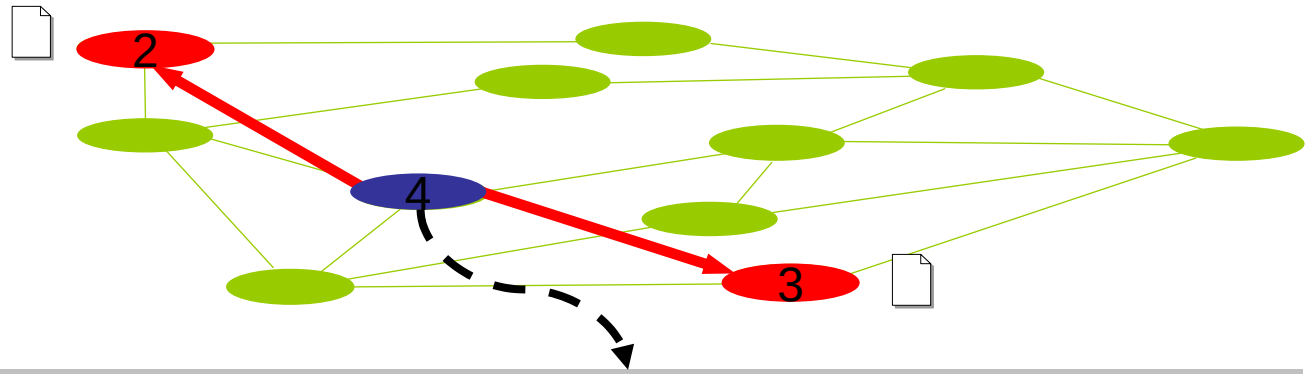
1. „Flood“ for existing Content Provider Short Cuts
2. Ask remote peer for Content Provider Short Cuts



Short Cuts Peer 1			
Query Q'	**PID**	**Query Hits**	**SC Type**
/SemanticWeb/RDF	2	100	C
/SemanticWeb/OWL	3	93	C

Create Recommender Shortcut Index (Active)

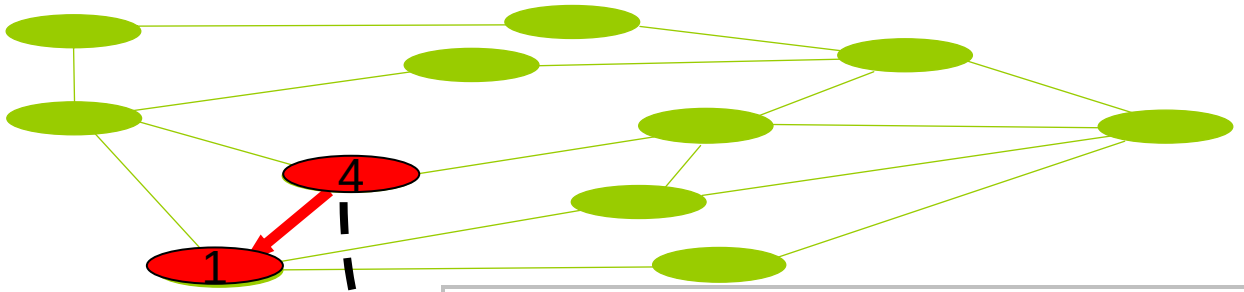
1. „Flood“ for existing Content Provider Short Cuts
2. Search Remote Peer for Content Provider Short Cuts
3. Learn Content Provider Short Cuts



Short Cuts Peer 4			
Query Q'	**PID**	**Query Hits**	**SC Type**
/SemanticWeb/RDF	2	100	C
/SemanticWeb/RDF	3	93	C

Create Recommender Shortcut Index (Active)

1. „Flood“ for existing Content Provider Short Cuts
2. Search Remote Peer for Content Provider Short Cuts
3. Learn Content Provider Short Cuts
4. Learn Recommender Short Cuts



Short Cuts Peer 4			
Query Q'	**PID**	**Query Hits**	**SC Type**
/SemanticWeb/RDF	2	100	C
/SemanticWeb/RDF	3	93	C
/SemanticWeb/RDF	1	193	R

Query independent shortcut

$$P.Bo = |shortcuts| \times |peers|$$

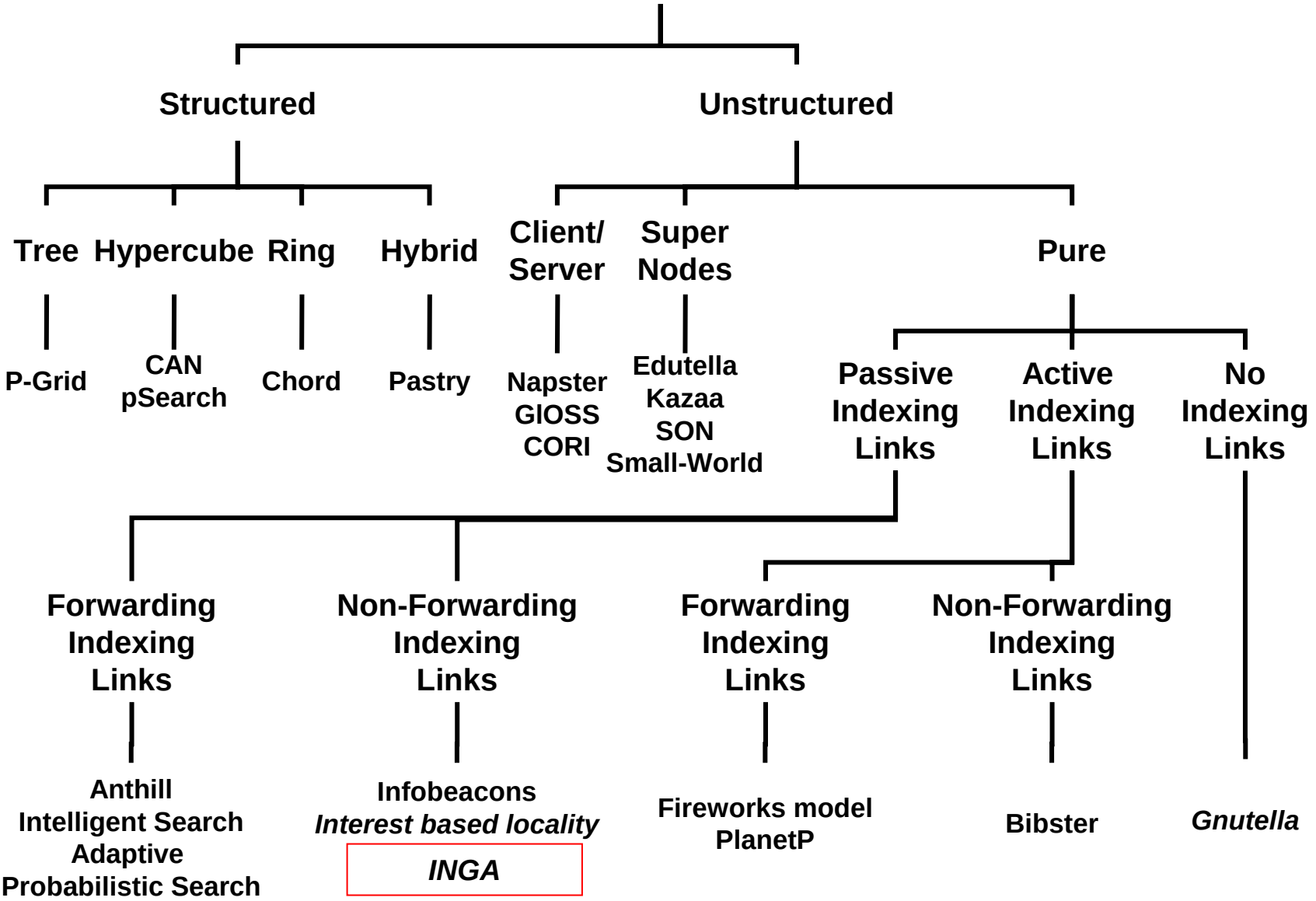
PID	Shortcuts	Peers	P.Bo
1	3	3	9

Query Independent Shortcut Index of Peer

Limit index size

- Retain only a small number of shortcuts in the index (e.g. 40 in our experiments)
- Delete based on combination of
  - Last update
  - Similarity to local content
  - Content vs. Recommender

Peer-to-Peer network organization



Simulation Environment

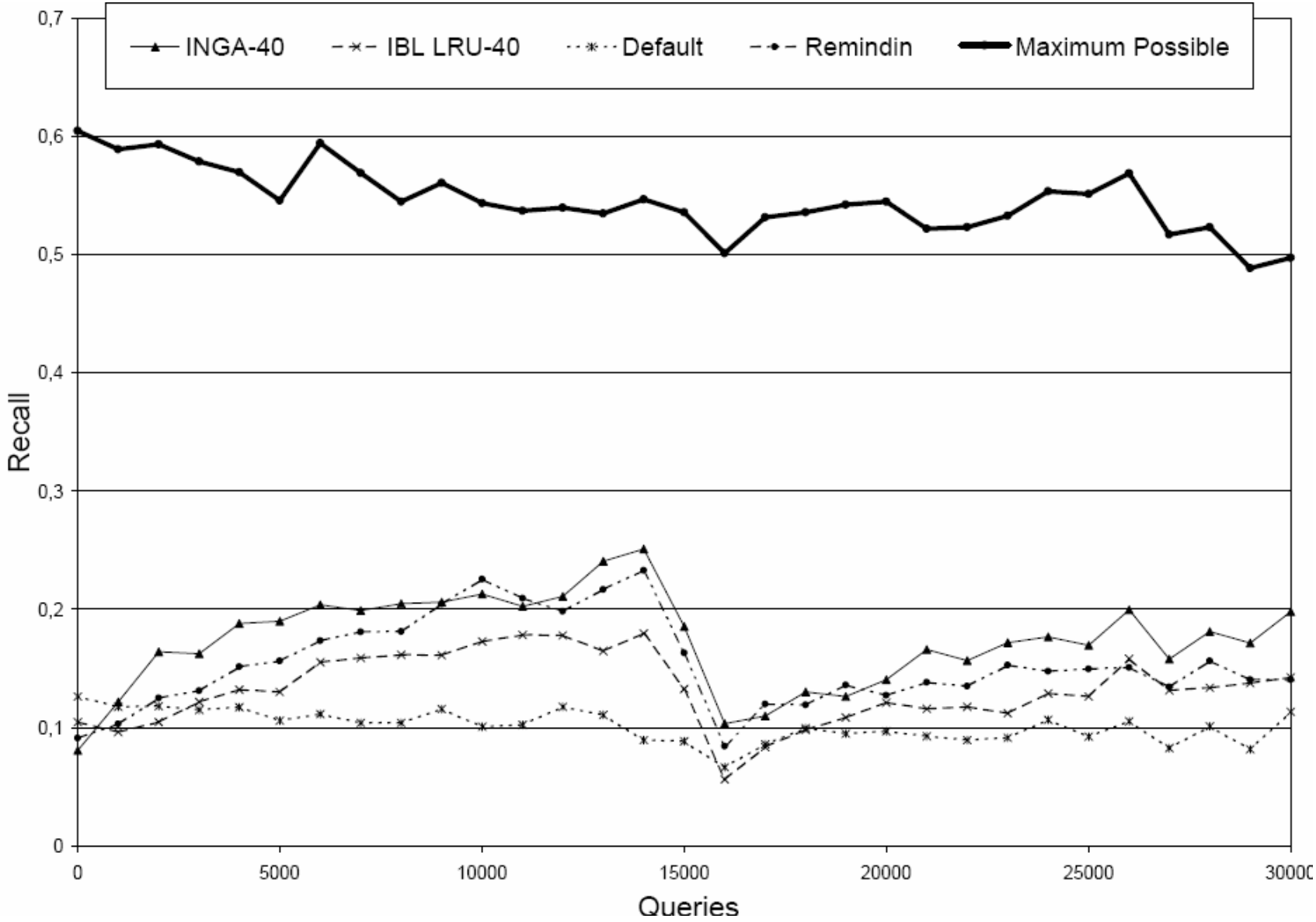
- Data Basis: Open Directory Project
- Content distribution: one editor = one peer
  - 1024 Peers, 1657 Topics
- Query distribution:
  - over topics: zipf
  - over peers: uniform
  - 30000 Queries, each peer in avg. 30 queries
  - 15000 Queries = one hour network time
- Network Topology: Power Law
- Top k=2, Max Hops h=6

- Goal: Compare against Gnutella and Interest-based Locality (IBL) approach

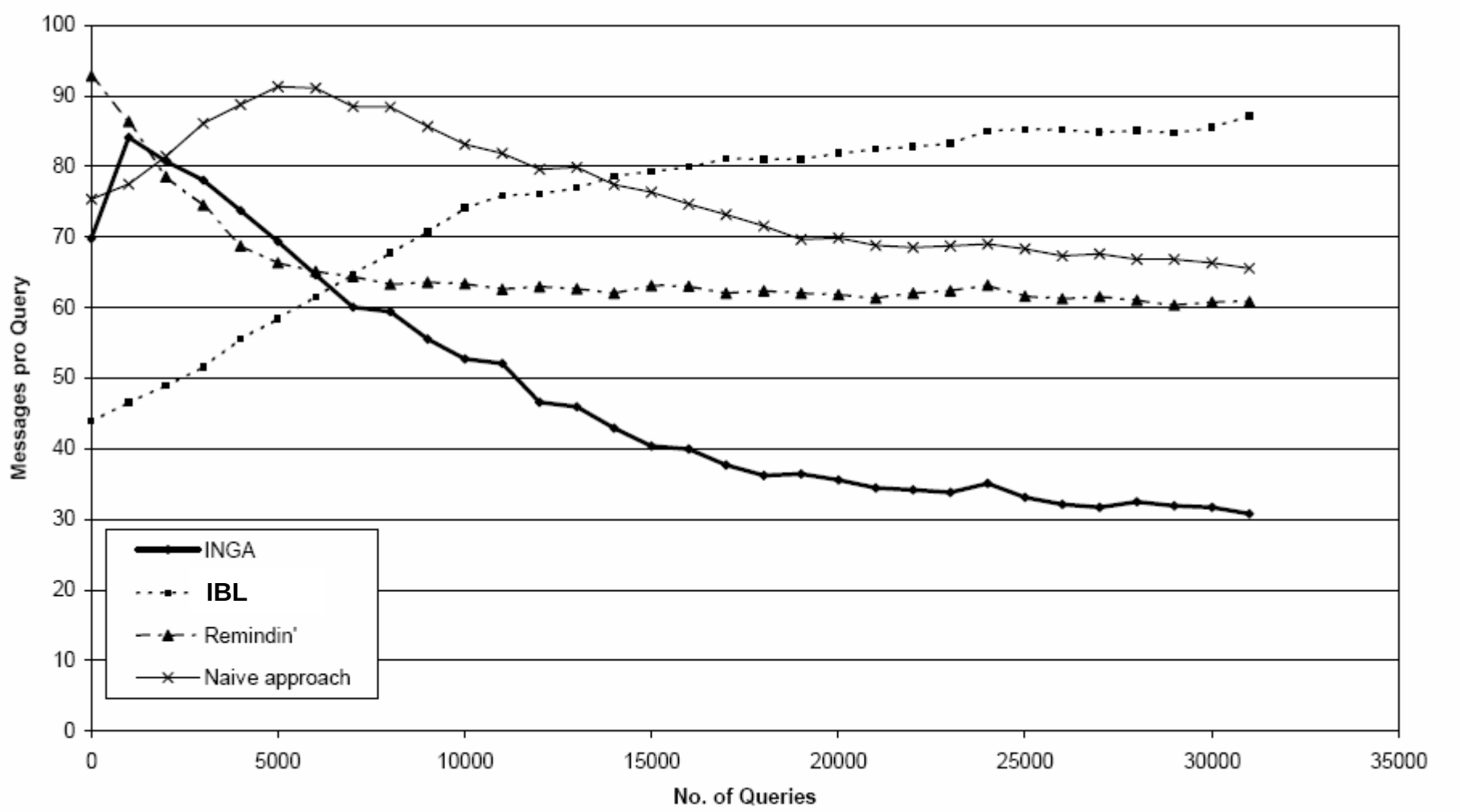
Experimental hypotheses

1. The algorithm performs at least equal in terms of recall than the naive algorithm and IBL (Sripanidkulchai et al.)
2. The algorithm performs better in terms of messages per query the naive algorithm and IBL.
3. The gain in efficiency can be attributed to equal account the different layers
4. A dynamic combination of query dependent and independent search strategies reduces the number of consumed per query while it retains a high recall.

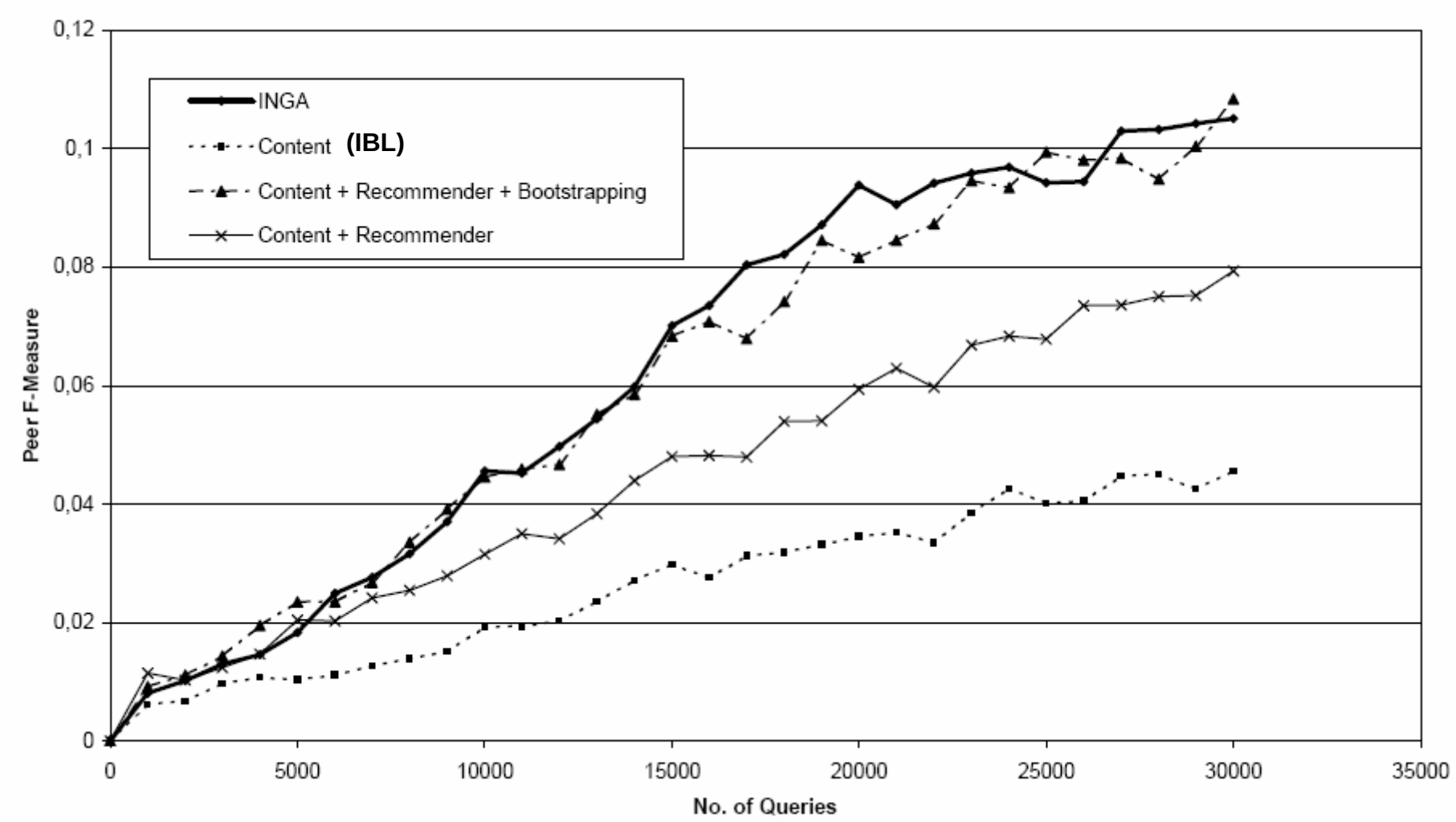
Effectiveness of Query Routing (recall)



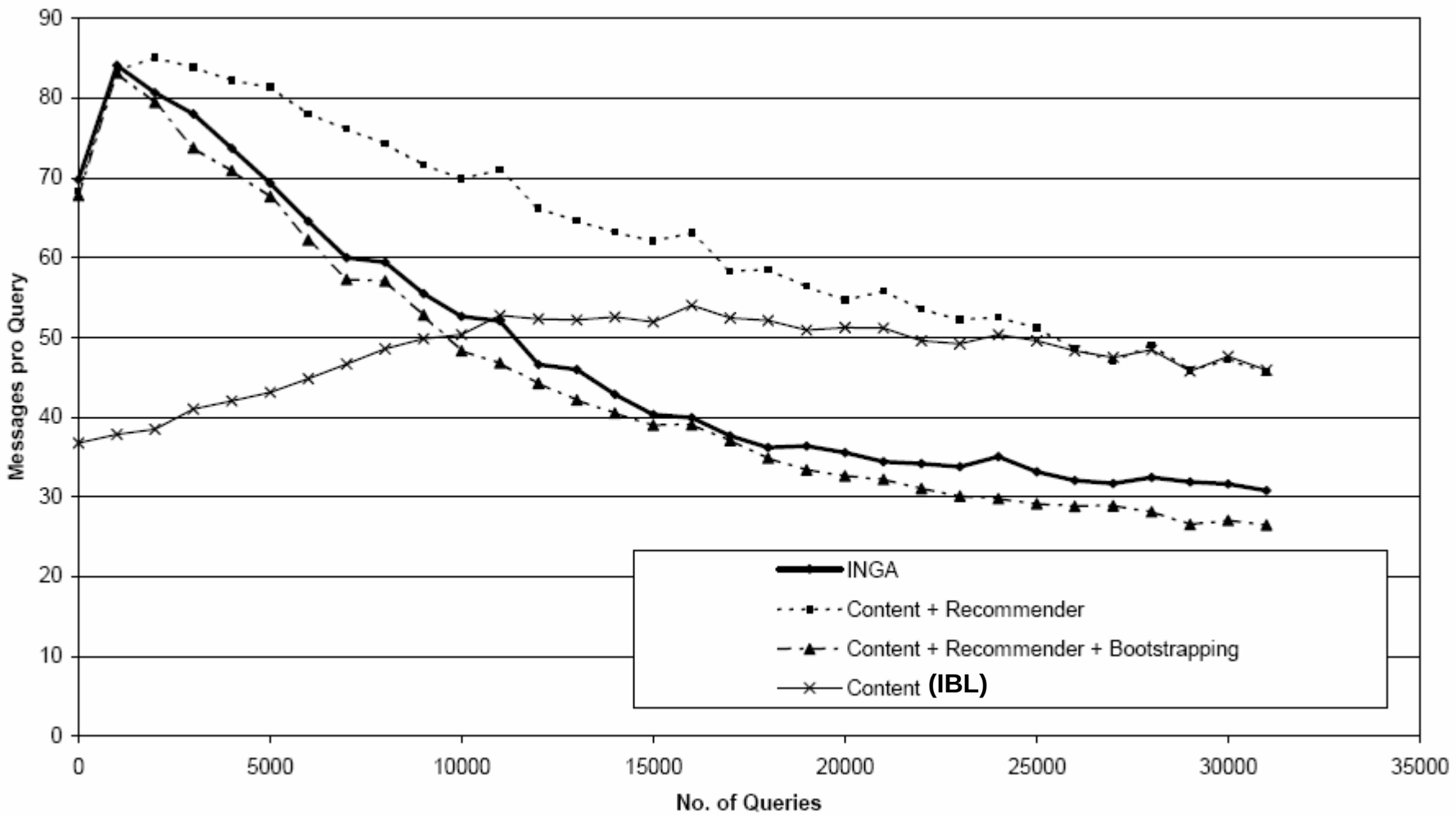
Efficiency of Query Routing (# messages)



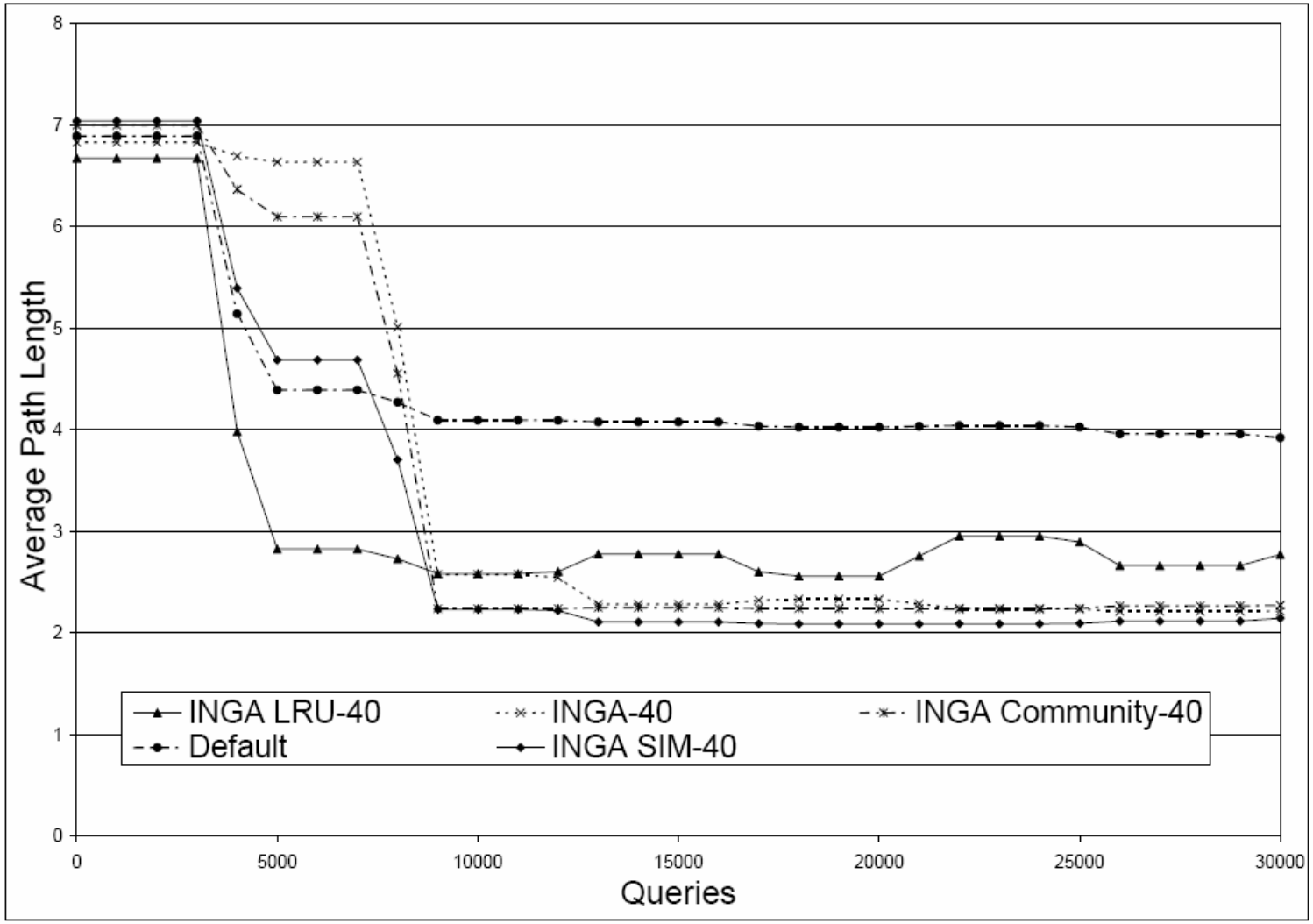
Contribution of Layers (peer f-measure)



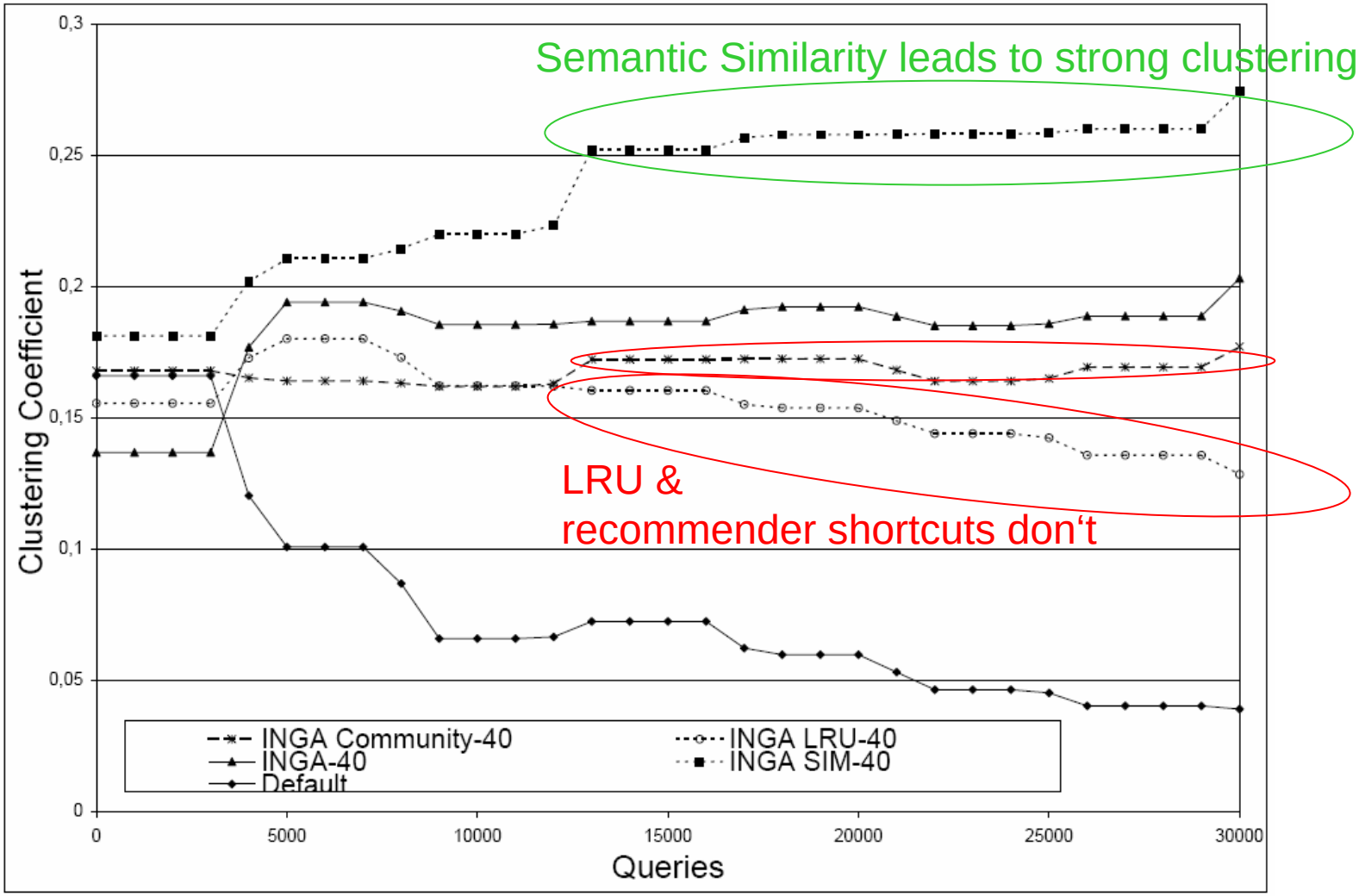
Contribution of Layers (# messages)



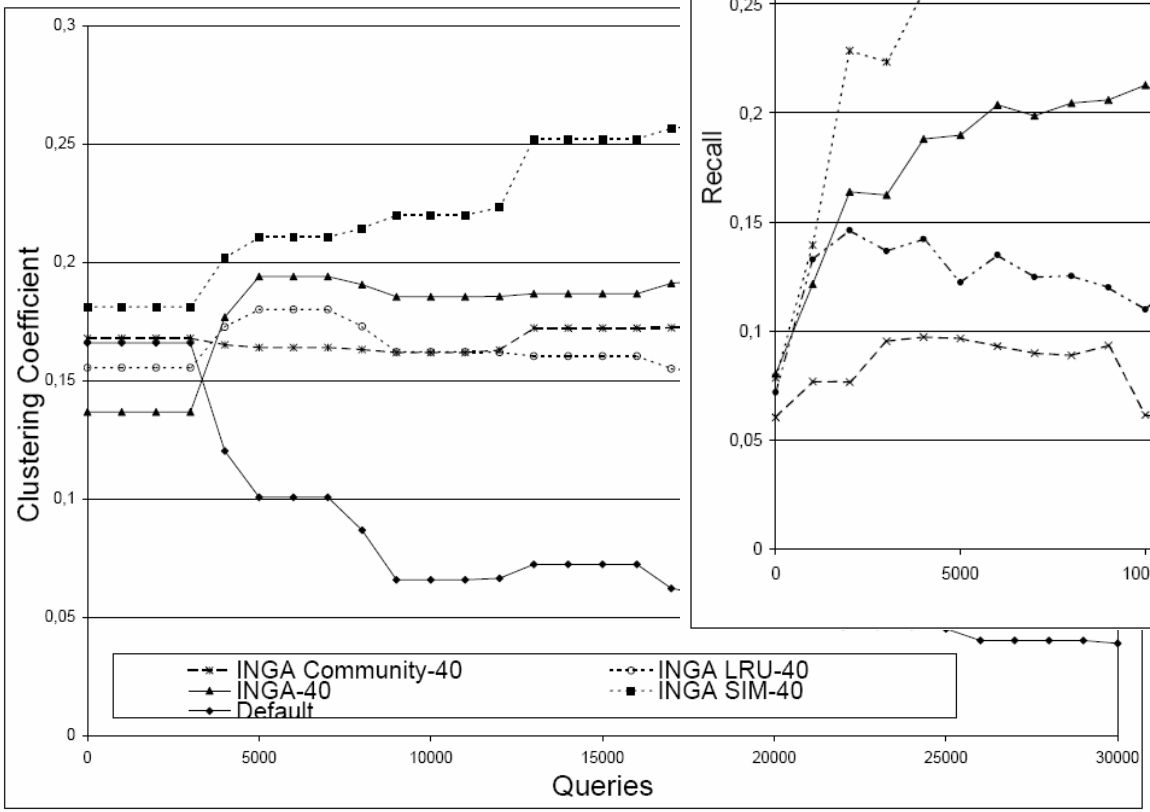
It‘s a Small World with Short Paths

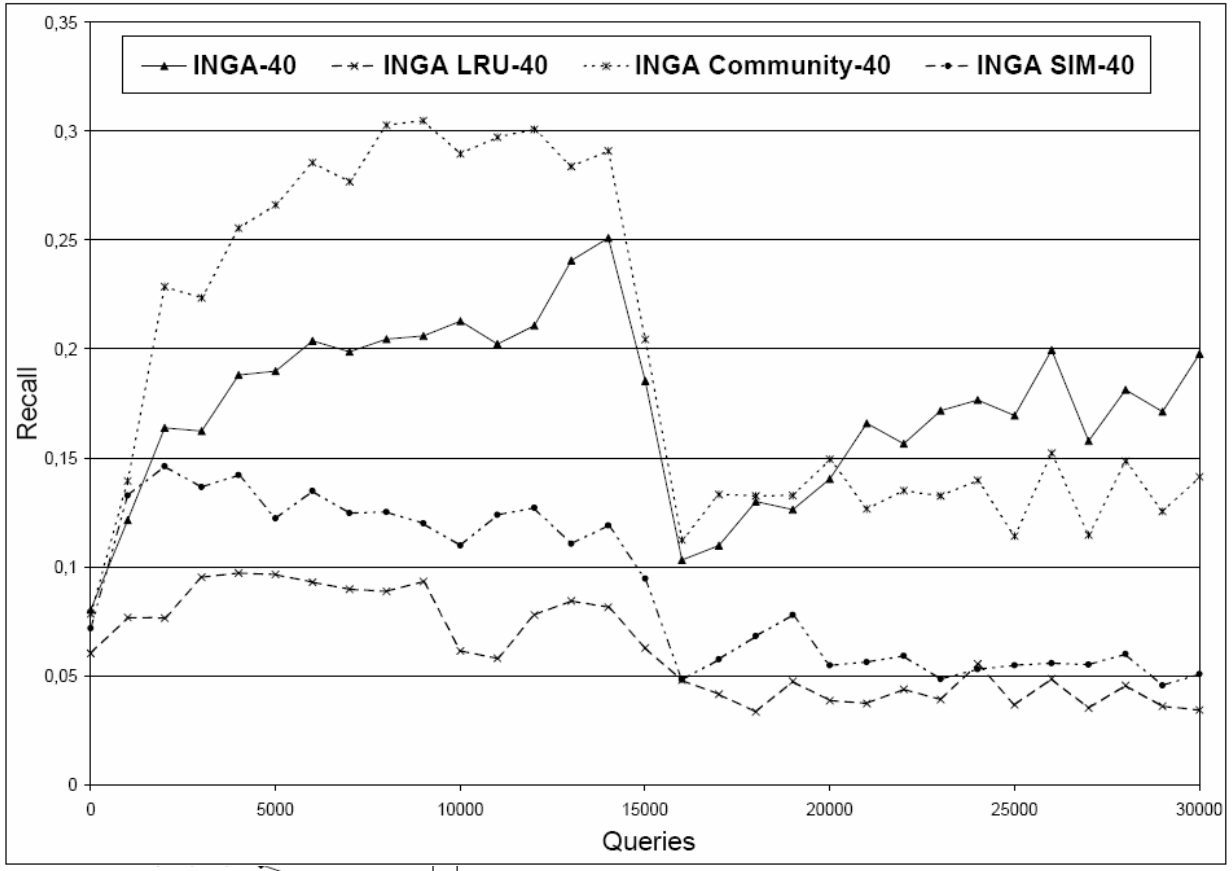


It‘s a Small World of Natural Clusters



…but, don‘t get trapped…

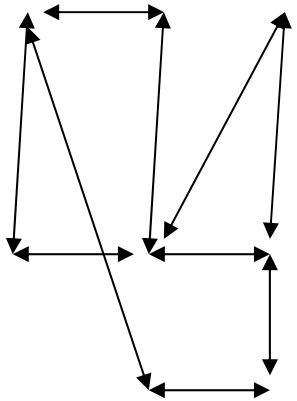




Social Networks Construction Challenge - revisited

Construction

- Nodes/Agents
- Construct Links
- Live in time
- Talk about topics
- Influence others

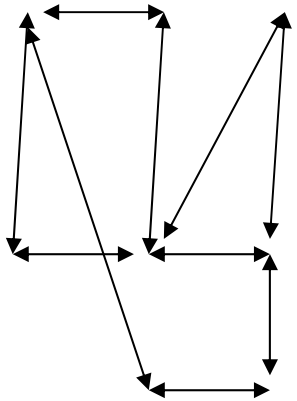


- Peers
- Query forwarding
- LRU
- Interest-based locality
- Index update content / recommender / bootstrapping

Social Networks Construction Challenge - revisited

Construction

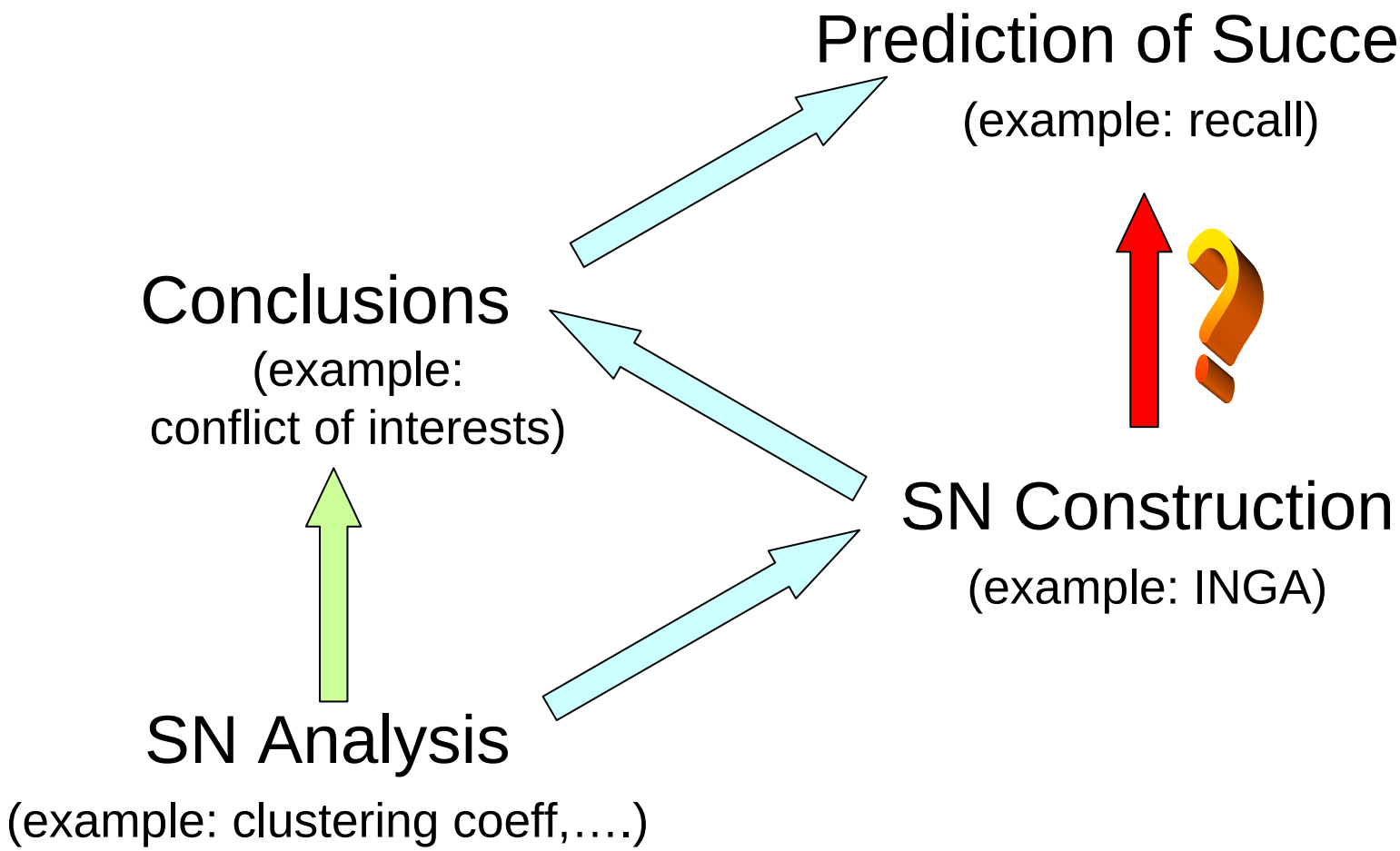
- Nodes/Agents
- Construct Links
- Live in time
- Talk about topics
- Influence others



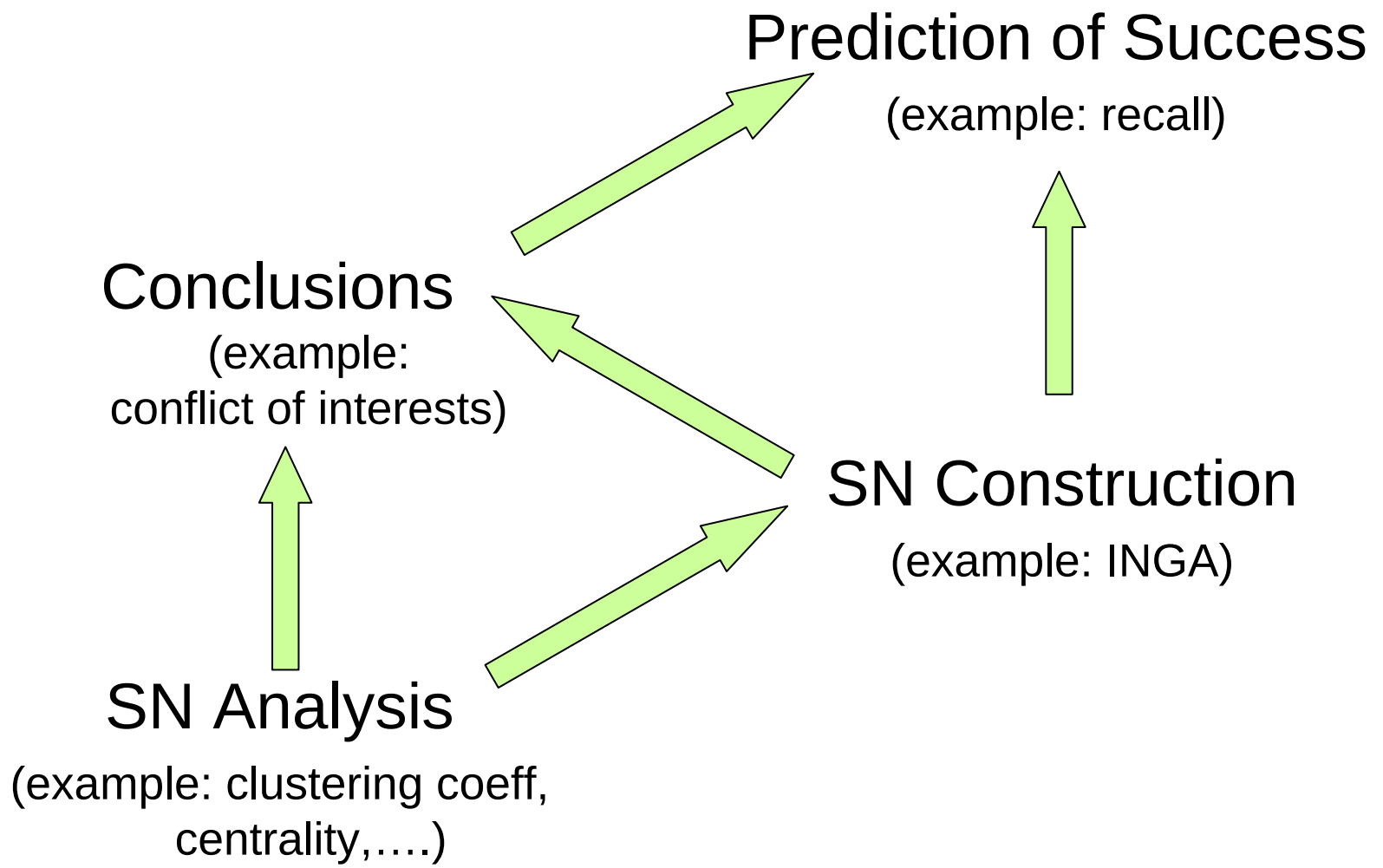
Success Criteria

- Effectiveness - recall
- Efficiency - #messages
- Robustness – reaction to change

Lessons learned, and some challenge for the future



Lessons learned, and some challenge for the future



Be careful: Advertisements instead of „Thank You!“

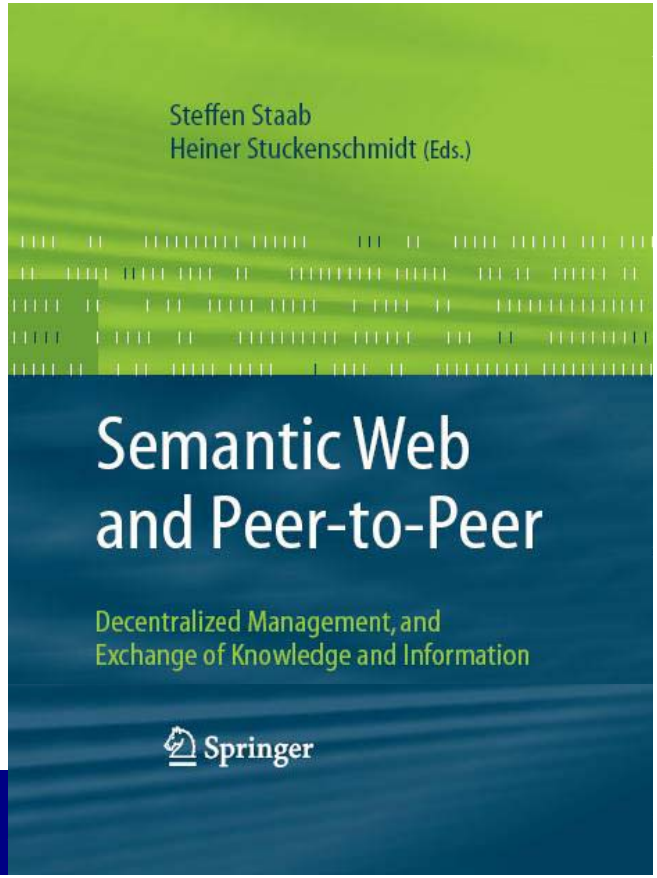
EKAW-2006 –
15th Conf. on Knowledge Engineering and Management

Managing Knowledge in a World of Networks,

http://ekaw.vse.cz/

Poster deadline, June 22, 2006

cf. A. Löser, S. Staab, and C. Tempich,
Semantic Social Overlay Networks,
IEEE JSAC –
Journal of Selected Areas in Communication,
to appear 2006/2007



Lessons Learned

Observation

- Huge hype around analysing social network (e.g. WWW)
- Comparatively little work about exploiting social network properties in a constructive way

I speculate:

- Social network properties may be used to improve networks of peers/different kinds
- Maybe: use richness of different network measures to optimize desired properties:
  - Here (?): Recall good if
    - High clustering co-efficient
    - Many long-distance links
    - ....