

Searching for “Familiar Strangers” on Blogosphere: Problems and Challenges

N. Agarwal, H. Liu, J. Salerno, and P.
Yu

<http://www.public.asu.edu/~huanliu/papers/ngdm07.pdf>

An Ever Evolving Field

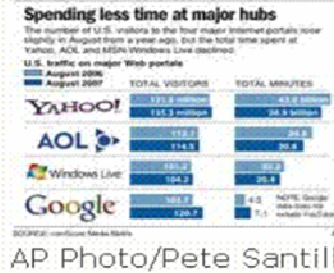
Oct 8, 7:48 AM EDT

Major Internet hubs see lesser influence

By ANICK JESDANUN
AP Internet Writer

NEW YORK (AP) -- The recent rush by major Internet portals to buy advertising companies and extend their sales networks is a sign that the business of being a one-stop shop for information and entertainment isn't what it used to be.

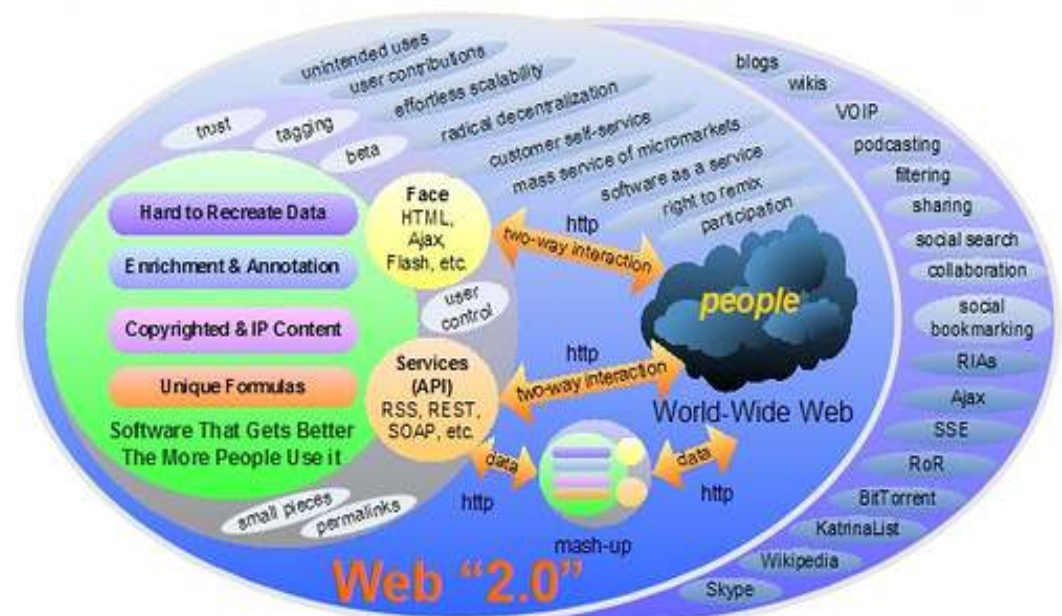
Gone are the days of emphasizing ways to attract and keep visitors - the way television



- Fragmentation alters Internet and presents new challenges to Next Generation Data Mining
- We present a problem encountered in our study
- The problem is about
 - Web 2.0
 - Blogosphere
 - Long Tail
 - Connecting dots
 - Challenges

Web 2.0

- Universal information space
- People generate and enrich contents
- Low barriers for publishing and participating
- Information consumers are now also Producers, or Prosumers
- Even more people are joining for various reasons from expression to reputation

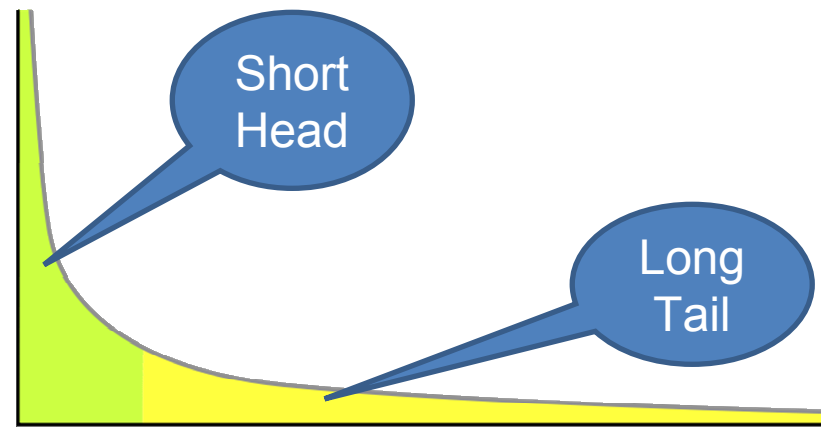


Blogosphere

- Blogosphere
- Blog site
- Blogger
- Blog post and comments (blog entries)
- Reverse chronologically ordered entries
- Single authored (Individual Blog Sites)
- Multi authored (Community blog sites)

A Growing Revelation: Short Head and Long Tail

- Few people are densely connected: Short Head
- Many people are sparsely connected: Long Tail
- Long Tail and Short Head tend to be equally important
 - Examples of albums *
- Zipf's Law, Power Law, Law of the Vital Few are well-known examples
 - The 80/20 rule becomes more skewed



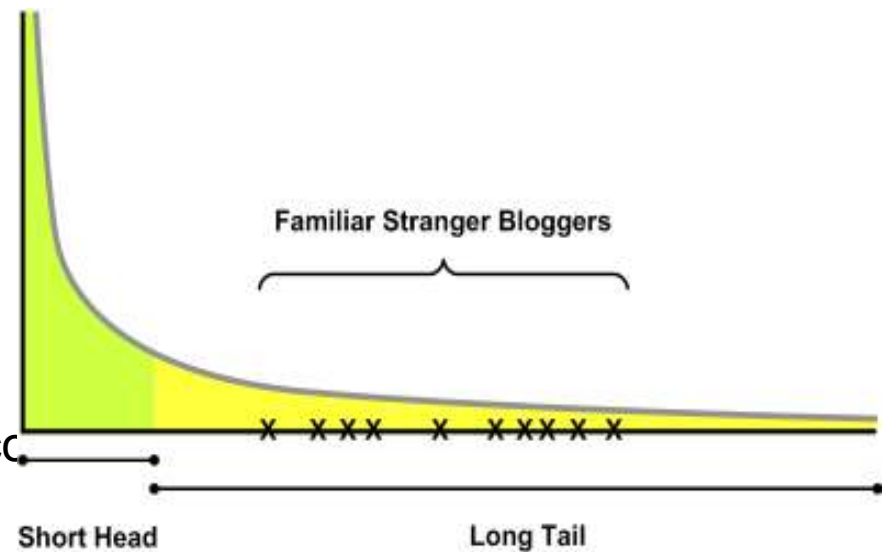
* The Long Tail, C. Anderson

Who are Familiar Strangers?

- People who do not know each other but share some common patterns
- Real World
 - Individuals can notice the presence of some others at a train platform
 - Common pattern: leaving at the same time and place
- Blogosphere
 - What you write is what you are ...
 - Have similar blogging behavior, interests (e.g., movies, games, technology, and politics, etc.)
 - Never cited (came across) each other

Bloggers in Long Tail

- Not returned as top hits by search engines
- Inordinately many
- Disconnected
- A Movie Review example
 - Critics – Short Head (e.g., nytimes.cc)
 - Movie Bloggers – Long Tail



- Long Tail is the most lucrative test-bed for finding **Familiar Strangers**

Finding Research Groups across Disciplines with Similar Interest

- An example of `feature selection' research
 - Many groups in different disciplines are working on that without calling it feature selection
 - Some working on Feature Selection but mean different things
- Web 2.0 Marketing 4Ps
- Numerous book/product reviews
- Another Web 2.0 application
 - Finding who are the best matches you don't know

Web 2.0 Marketing 4Ps

- Personalization for the long tail
 - Much easier to fish where the fish are
- Participation
 - Creating a critical mass and generating crowd effect
- Peer-to-peer
 - Online word of mouth
 - Smooth information flow
- Predictive modeling
 - Anticipation and recommendation

Connecting the Dots

An example from The Long Tail book

- Touching the Void, 1988
 - Largely forgotten after its publication
- Into Thin Air, a decade later
 - Its publication attracted *a few readers* who wrote reviews that connect this book to Touching the Void
 - Touching the Void became a new sensation

Aggregating Niches in Long Tail

- A blogger with familiar-strangers will form a critical mass such that
 - the understanding of one blogger gives us a sensible and representative glimpse to others,
 - more data about familiar strangers can be collected for better customization and services (e.g., personalization and recommendation),
 - the nuances among them present new business opportunities, and
 - knowledge about them can facilitate predictive modeling and trend analysis.

Need for Aggregation

- Customized attention requires substantial data
- Majority of blog sites are in the Long Tail
- ...and are disconnected
- Aggregating the similar yet disconnected for obtaining critical mass
- Lack of data can result in irrelevant ads (see an example on the right)
- Increase participation
- Move from the Long Tail closer to the Short Head
- Smooth knowledge transfer between familiar strangers

[Uncle Walt says the new iMac rocks Vista](#)

Posted Aug 25th 2007 7:00PM by [Mat Lu](#)

Filed under: [OS](#), [Switchers](#), [iMac](#)

Ever since [Boot Camp](#) was released it has been no surprise to find out that the Intel Macs also make for some of the best Windows machines too (well, if you can bring yourself to install it, that is).

Anyway, Walt Mossberg, dean of tech writers, has gotten one of the new iMacs and for kicks installed Vista via Boot Camp. And sure enough, [Uncle Walt says](#) he tested it "using Vista's built-in Windows

Experience Index, a rating system that goes from 1 to 5.9, with scores above 3.0 generally required for full, quick performance. My iMac scored a 5.0, the best score of any consumer Vista machine I have tested." This was apparently the 2.8GHz machine as he says it was the top-of-the-line model. I know some folks were [disappointed](#) with the new iMacs slightly anemic graphics cards, but it's good to know they can still rock Vista if called upon to do so.



[via [MacVolPlace](#)]

[Read](#)

[Permalink](#)

[Email this](#)

[Linking Blogs](#)

[Comments](#) [11]

[Electronic Paper](#)

PVI is the world's first flexible & only active matrix e-paper maker

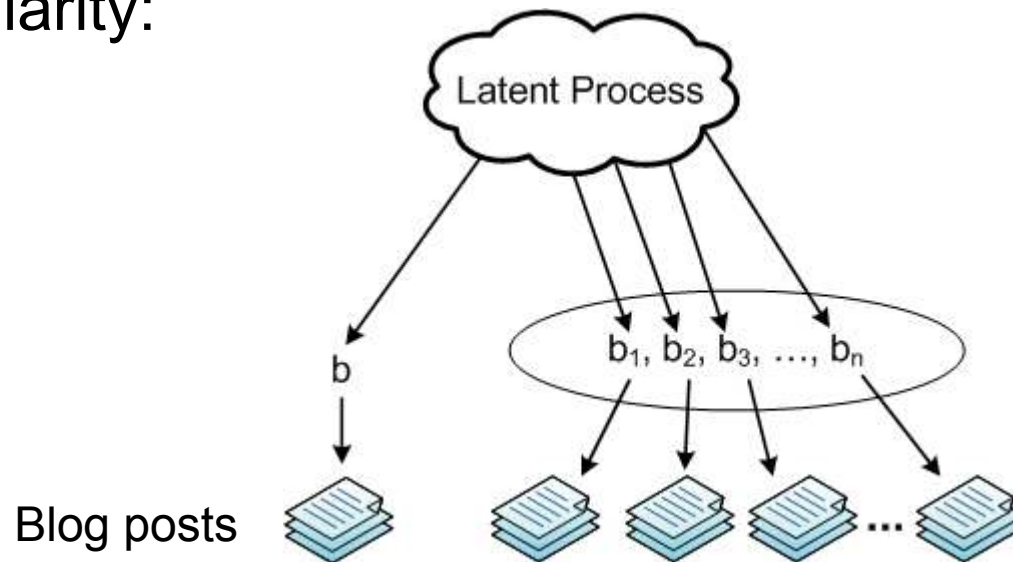
[Your Momma Blogs](#)

And we pay her to. You can get paid to blog and make money too.

Ads by Google

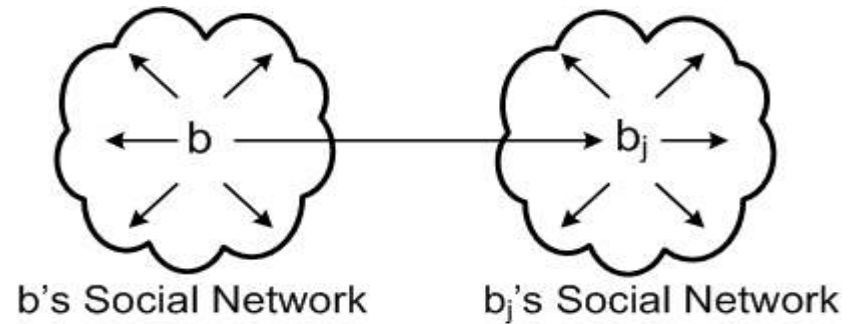
Definition

- Given a blogger b , **familiar strangers** to b are a set of bloggers $B = \{b_1, b_2, \dots, b_n\}$, who share common patterns as b , like blogging on similar topics, but have not come across each other or are not related to each other.
- Familiarity:

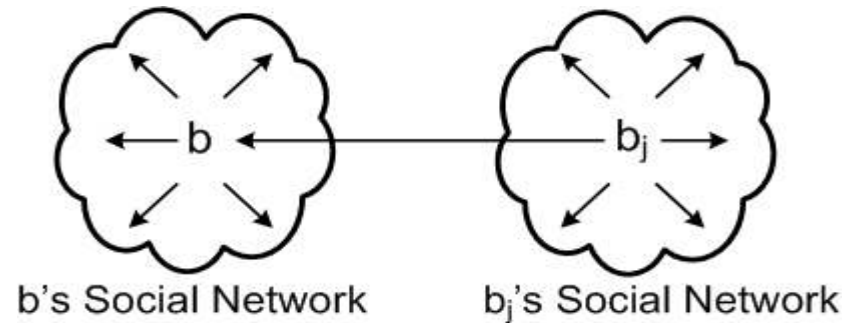
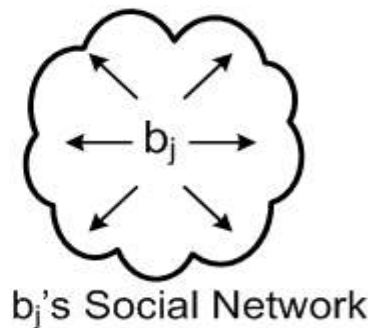
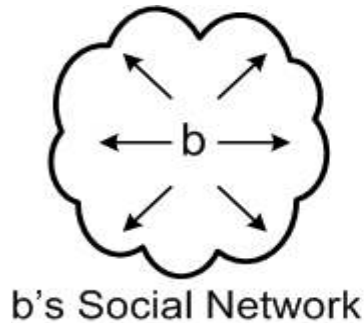


Definition

- Strangers:
 - Partial strangers
 - Total strangers



b_j is in b's Social Network



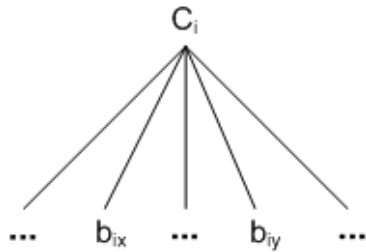
b and b_j have disjoint Social Networks

b is in b_j's Social Network

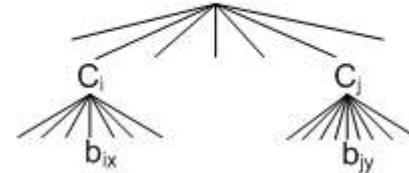
Different Types

- Organizational differences in the blogosphere eventuate disparate types of familiar stranger bloggers

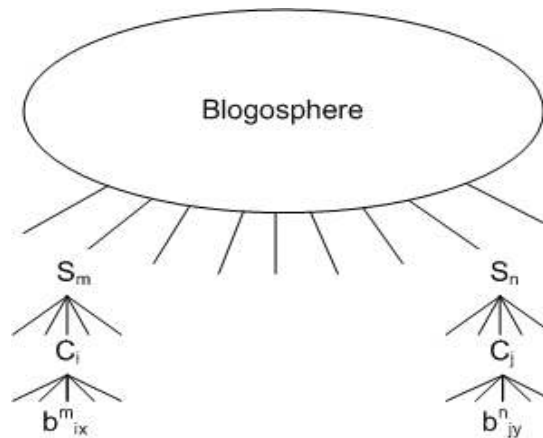
Community-level
familiar strangers



Networking-site-
level familiar
strangers

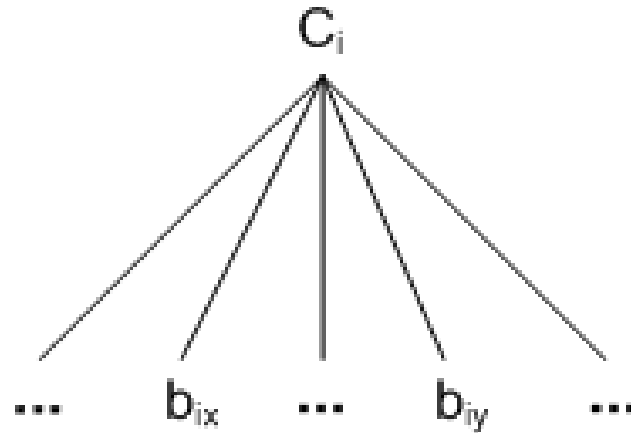


Blogosphere-level
familiar strangers



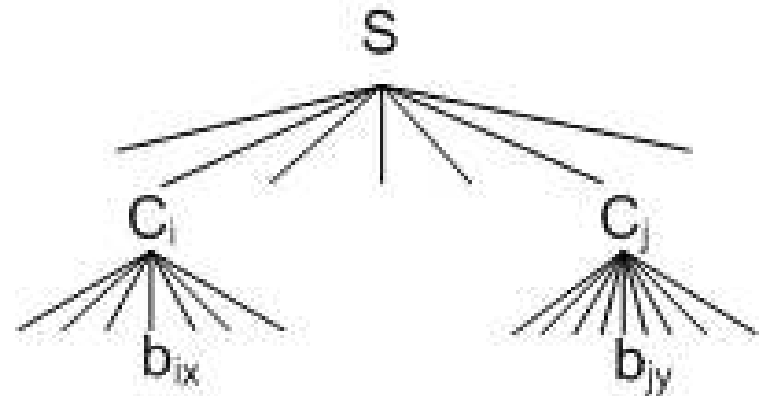
Community-Level Familiar Strangers

- MySpace has a community called “A group for those who love history”
- It has 38 members
- Two members, “Maria” and “John”
 - blog profusely on the similar topic,
 - but they are not in each other’s social network



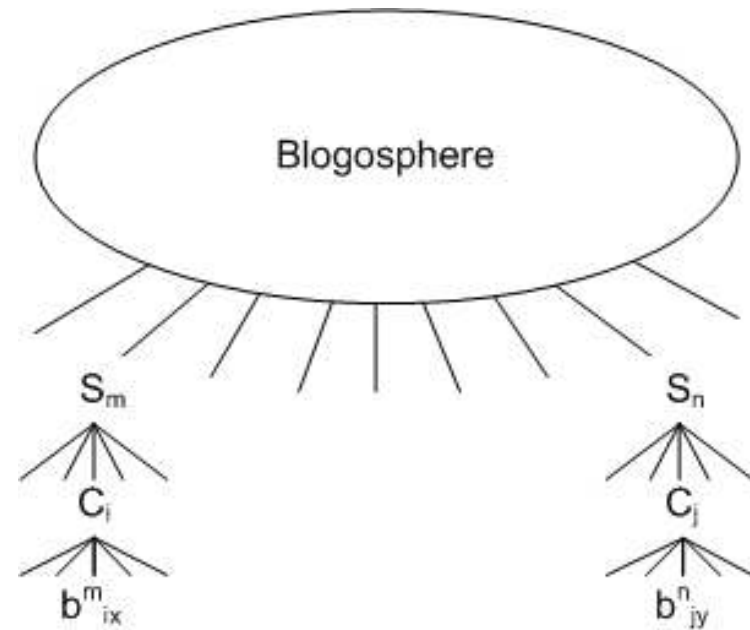
Networking-Site-Level Familiar Strangers

- 2 groups on **MySpace**,
 - **The Samurai** (32 members)
 - **The Japanese Sword** (84 members)
 - **Marc**, top blogger on “The Samurai” and **Jeff**, top blogger on “The Japanese Sword” discuss about **Japanese martial arts**.
 - Neither of them is in the other’s social network.
 - This implies, though being active locally and discussing on the same theme, the two bloggers are still strangers.



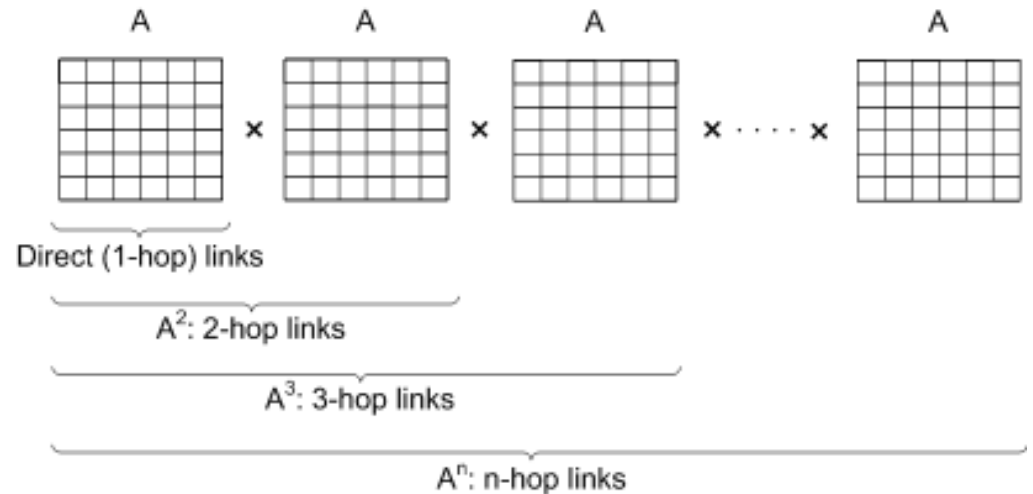
Blogosphere-Level Familiar Strangers

- 2 different social networking sites, **MySpace** and **Orkut**.
 - **The Samurai** (32 members) from MySpace
 - **Samurai Sword** (29 members) from Orkut
 - Top bloggers from the respective communities in MySpace and Orkut, **Marc** and **Anant**, respectively, share the blogging theme but they are not in each others' social network.
 - The above example illustrates the existence of blogosphere-level familiar strangers.



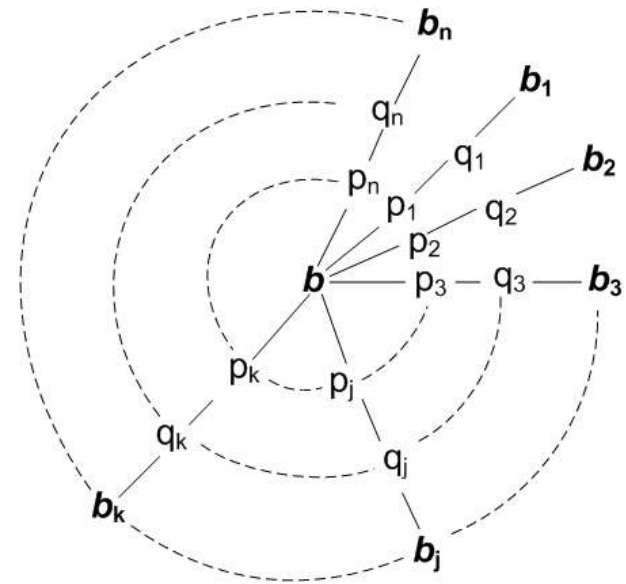
Challenge

- Link analysis
 - Exhaustive search is impractical
 - One step-look ahead cannot go too far



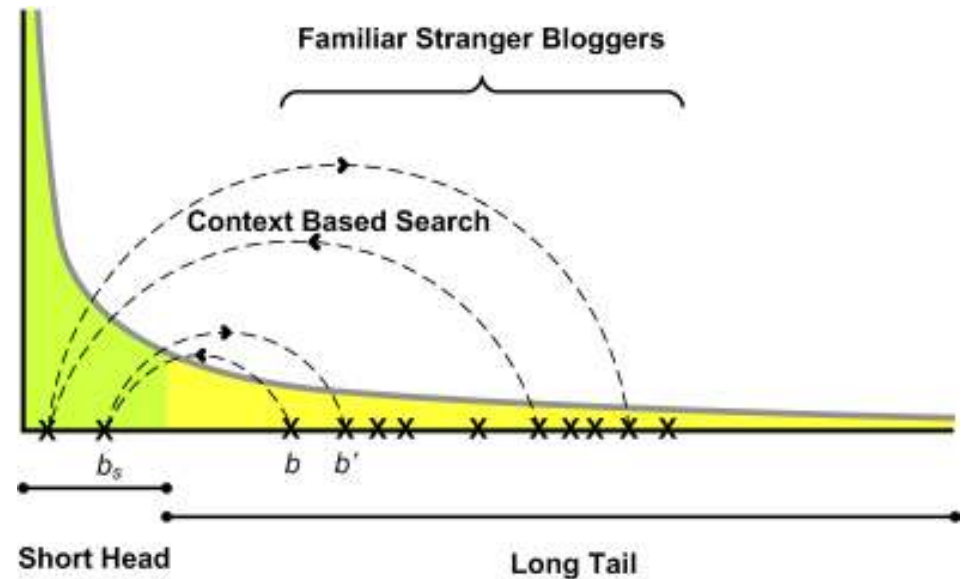
Challenge

- Searching via blog posts
 - Defining Similarity
 - If we know exactly what we want, ...
 - For the example of two books (TTV, ITA), ...



Challenge

- Using the context to help search
 - What constitutes the context for a blogger
 - Directed vs. broadcast search



An intuitive approach is to integrate link-, similarity-, and context-based methods

Yet Another Challenge

- How to evaluate the work without “standard” ground truth?
 - Data and data collection
 - Experiments
 - Performance metrics

Ending Remarks

- Finding “familiar strangers” on the blogosphere is challenging
- Aggregating familiar strangers can help find niches in the Long Tail
- We present more problems than solutions
- Long Tail is where abundant new opportunities are

- A related note: Workshop on [Social Computing, Behavioral Modeling, Prediction](#)

<http://www.public.asu.edu/~huanliu/sbp08>