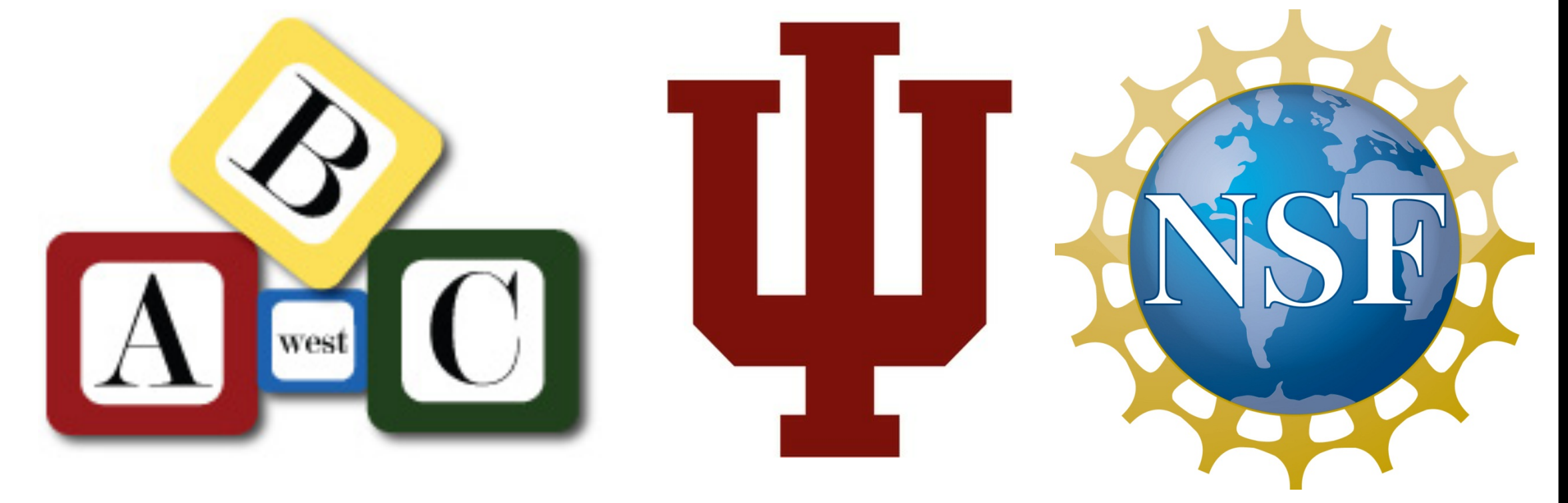


# Social information environments of collaborative tagging systems: Individual and group-level cognitive perspectives

Jared J. Lorince, Saurabh Malviya, and Peter M. Todd



## Background and Motivation

The proliferation of the Web and other related technologies has driven humans to adapt to new kinds of **social information environments**, utilizing cognitive mechanisms that originally evolved for search and decision-making in physical spaces. This work aims to increase our understanding of how these environments both **constrain** and are **shaped** by the cognitive processes of their users.

Collaborative tagging systems (e.g. Last.fm, Delicious.com, Flickr.com) represent interesting domains in which to study these phenomena, as they allow us to examine the dynamics of an information environment that is simultaneously the product and a determinant of users' behavior. We describe a research program for studying the information environments of such systems.

## Principal research questions

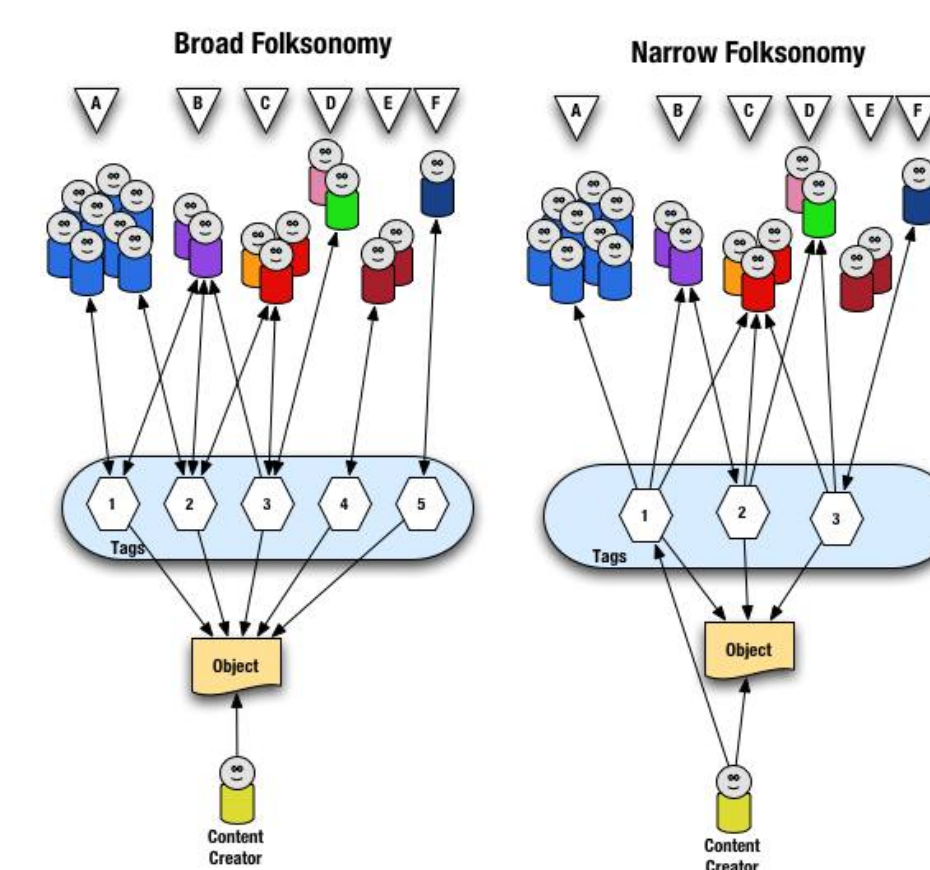
1. What **motivates** people to participate in collaborative tagging systems?
2. What **heuristics** and other decision-making strategies govern their tagging choices?
3. How do these decisions manifest at the aggregate level, and how does the **evolving choice environment** in turn affect users' decision making?

We will approach these questions with a focus on the social music site **Last.fm**, on which users can tag artists, albums, and songs.



## Motivations & Heuristics

**Why** do people utilize collaborative tagging systems in the first place? Researchers have addressed this question using surveys [4], and distinguish between **broad** and **narrow** folksonomies [5]. Preliminary work exploring users' **tagging styles** suggests broad variation in how people utilize tagging systems, and understanding their differing motivations will be critical to our research program



Schematic view contrasting narrow (in which people mainly tag their own content) and broad (in which many unique people tag the same items) folksonomies.

We also expect **heuristic decision-making** to be common in tagging environments, in which:

- There are large quantities of information on which to **base tagging decisions**
- The number of **existing tags** to choose from is comparably large

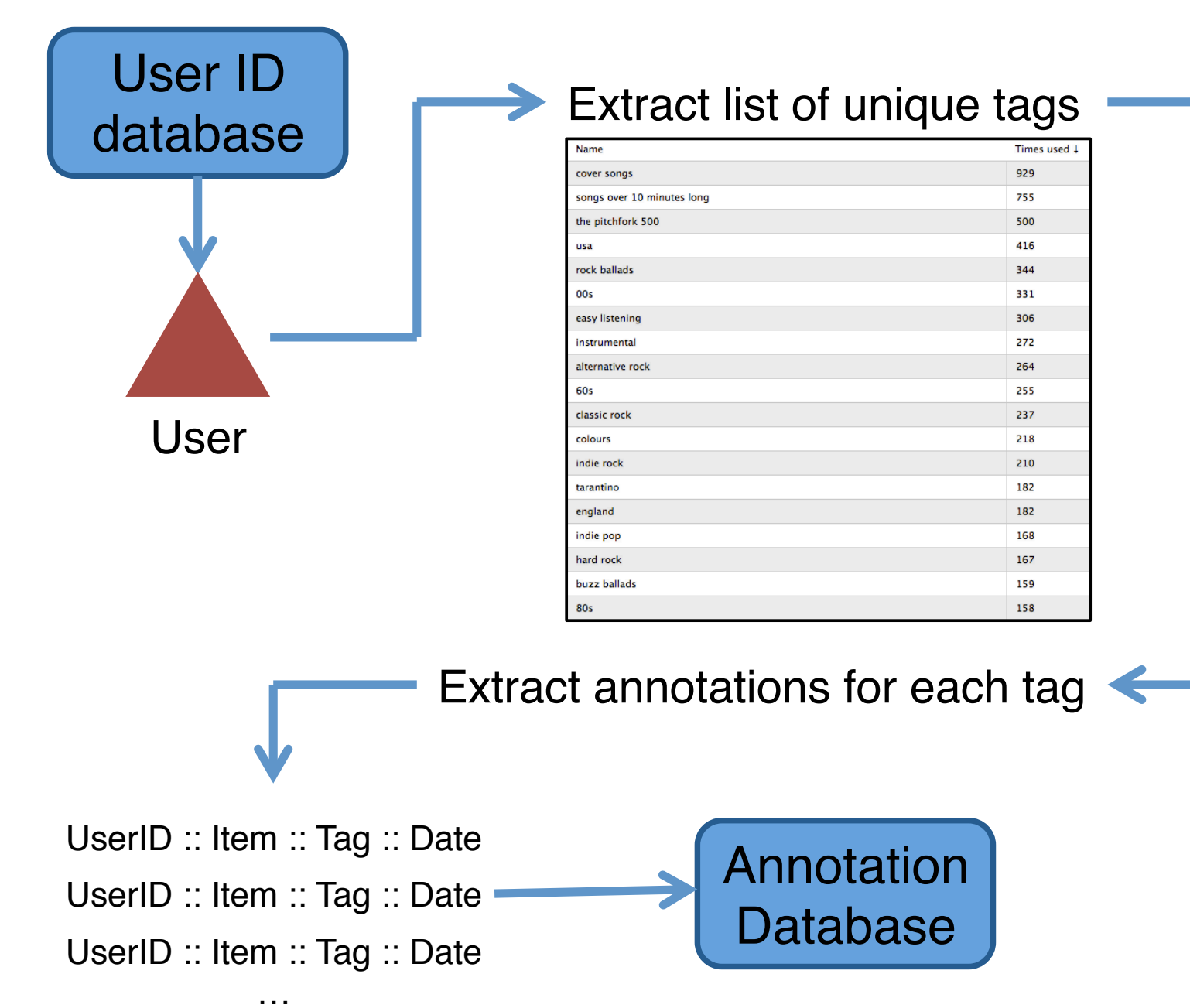
In particular, social copying strategies (matching the most popular tag(s) for a given item) are likely, and compel us to study how tagging decisions vary as a function of the (social-constructed) tag data displayed to users in the tagging interface.

## “Naturalistic” Tagging : Crawling Last.fm

The Last.fm web interface stores complete historical tagging information for all users. We have designed a **Web crawler** (data collection in progress) to extract tagging data for a large sample of users (100k+).

Track	Tags
progressive metal	progressive metal
Between the Buried and Me - Prequel To The Sequel	October 2011
Mastodon - The Sparrow	October 2011
Dream Theater - Take the Time	October 2011
Mastodon - Aqua Dementia	October 2011
The Hunter	September 2011
Symphony X - Iconoclast	June 2011
Mastodon - The Czar I, Utopia II, Escape III, Martyr IV, Spiral	May 2011
Gajira - The Way of All Flesh	May 2011
Between the Buried and Me - Augment of Rebirth	May 2011
Epic - Kingdom of Heaven (A New Age Dreams - Part V)	May 2011

Example tag history page. This displays a sample of the items this user annotated with the tag “progressive metal”



Crawler architecture. The process described is repeated for all sampled users

This data will let us analyze questions such as:

- How the distribution of tags assigned to specific content evolves over time
- How individual users' tagging tendencies change over time
- How the use of specific tags changes over time

Though this data will be useful for high-level analysis to address question #3 above, and may give some clues as to questions #1 and #2, to understand the decision-making heuristics driving individuals' tagging behavior, we need data at a **finer temporal resolution**, and the ability to **systematically manipulate the information environment** provided to users. This motivates the second major component of our research program.

## Tagging in the Lab: An Experimentation Platform

We are also developing a custom **collaborative tagging research platform**. It will mimic the functionality of Last.fm, allowing study participants to tag content and view the tagging behavior of other users. Further, we will implement:

- Selective display of tag information (existing tag data, suggested tags, etc.) in the tagging and item information (e.g. tag cloud) interfaces
- High-resolution timescale recording of changes in tag distributions
- Detailed logging of users' interactions with the tagging interface

This functionality will offer highly granular data on how individual tagging decisions vary with changes in the information environment, providing insight into all three of our main research questions.



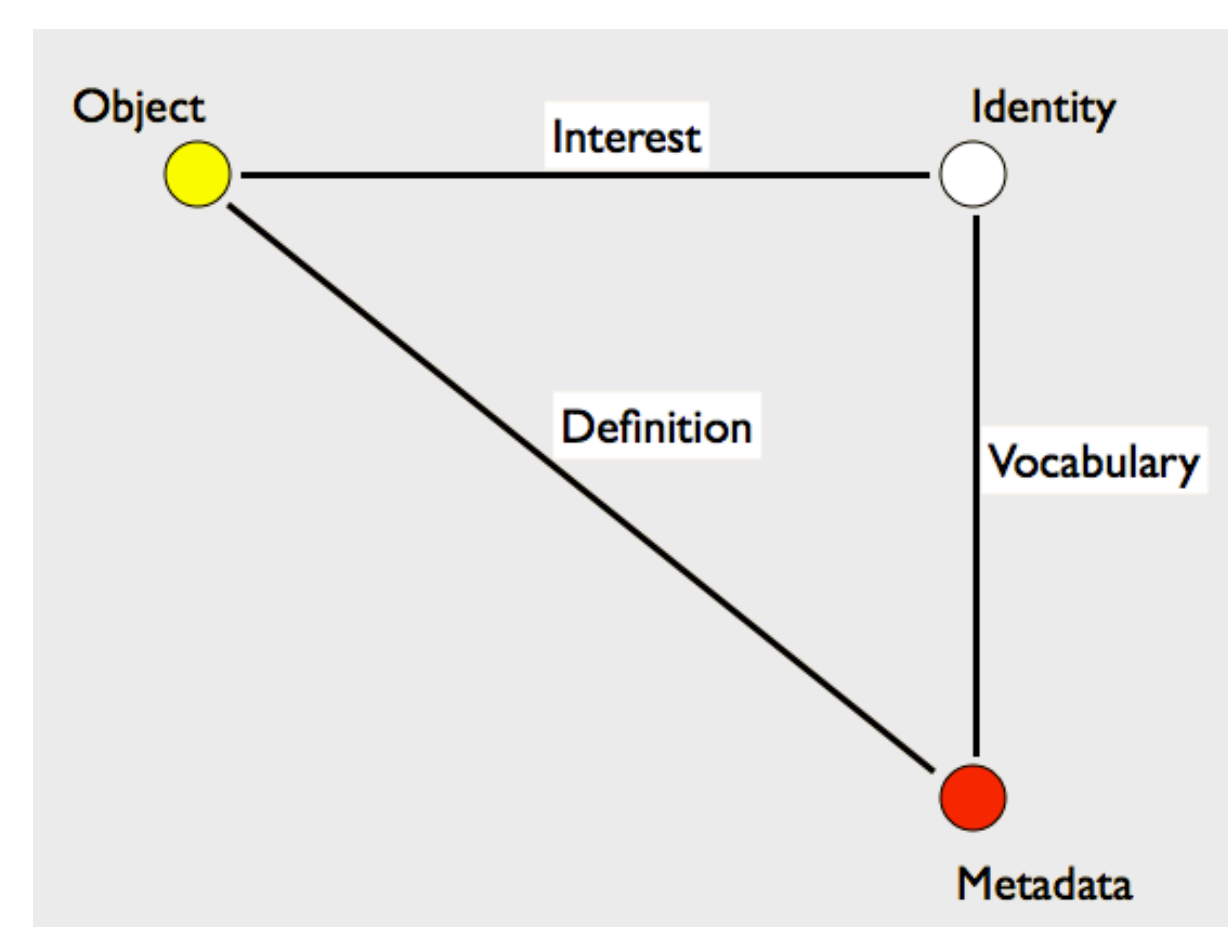
The Web-based tagging interface for Last.fm, upon which ours will be modeled.

## What is collaborative tagging?

Collaborative tagging systems allow users to annotate items (songs, bookmarked URLs, images, etc.) with **freeform text strings**. Typically, users can assign any number of unique tags to a given item (though no individual can assign the same tag to the same item more than once). Aggregating the annotations of many users, tagging systems generate a **“folksonomy”**, a crowdsourced categorization scheme for the items indexed by the service.

00s 90s acoustic **alternative** alternative rock amazing **ambient** art rock atmospheric  
avant-garde awesome beautiful chill chillout downtempo dream pop dreamy easy listening **electronic**  
electronica emo emotional epic ethereal **experimental** experimental rock favorites favourites folk Iceland  
**icelandic indie** indie pop indie rock instrumental love male vocalists melancholic melancholy  
mellow minimalist nordic pop **post-rock** progressive progressive rock  
psychedelic relax relaxing rock scandinavian shoegaze sigur ros singer-songwriter sleep slow-motion rock space  
rock trip-hop want to see live

A common visualization of tags is a so-called “tag cloud”. It displays the most popular tags for a given item (in this case the band Sigur Rós), with the font size of each tag proportional to the number of times it was assigned to the item



A folksonomy can be represented by a set of triads, each corresponding to an annotation and consisting of a user (“identity”), the item tagged (“object”), and the tag assigned to it (“metadata”). Image credit to Thomas Vander Wal (2006)

## Broader impacts

Our ultimate goal is to design collaborative tagging systems for various domains (e.g. academic paper databases) that help users more easily find **valuable** and **pertinent** information. This will require determining “good” tagging behavior, and then develop the means to encourage it using, e.g., **motivational game mechanics** (point systems that publicly reward users for the quantity and quality of their annotations, etc.)

## References

- [1] Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2), 198–208.
- [2] Schifanella, R., Barrat, A., Cattuto, C., Markines, B., & Menczer, F. (2010). Folks in folksonomies: social link prediction from shared metadata. *Proceedings of the third ACM international conference on Web search and data mining*, 271–280.
- [3] Nov, O., & Ye, C. (2010). Why do people tag? *Communications of the ACM*, motivations for photo tagging, 53(7), 128.
- [4] T. V. Wal. Explaining and showing broad and narrow folksonomies, 2005. [http://www.personalinfocloud.com/2005/02/explaining\\_and\\_.html](http://www.personalinfocloud.com/2005/02/explaining_and_.html).