

JOANNEUM



RESEARCH

The Wisdom in Tweetonomies

Acquiring Latent Conceptual Structures from Social Awareness Streams

Claudia Wagner

claudia.wagner@joanneum.at

Markus Strohmaier

markus.strohmaier@tugraz.at

a TRADITION of INNOVATION



Social Awareness Streams (SAS)

- Short, natural language messages created by users
- Broadcasted
- Information consumption is driven by social networks
- Applications such as Twitter or Facebook



marshalk Thought: if there's a black market 4 FB user data, then making academic researchers destroy their datasets is like prohibition of fine wine
4 minutes ago via Tweetie



petef Help stop the #DEBill being rushed through
<http://secure.38degrees.org.uk/stop-the-bill>
6 minutes ago via Tweetie



futureweb2010 tech companies seek stronger #privacy laws:
<http://tinyurl.com/yatq4t4> learn more about the future of privacy and the #web @futureweb2010
12 minutes ago via web



RWWRSS ReadWriteWeb: Where 2.0: Quova and SimpleGEO
Make App Building Easier <http://bit.ly/d6gcVQ>
14 minutes ago via twitterfeed



joshsh Tweaking a #twitlogic instance to capture stock market buzz. @xenobaby has some really cool ideas in this area. #twcrpi
17 minutes ago via TweetDeck



cygri Congrats! RT @dpalmisano: Wow! my presentation (<http://bit.ly/chsAdt>) on #linkeddata has been cited on readwriteweb <http://bit.ly/awrVxP>
18 minutes ago via Twitterrific

[Naaman, 2010]

.....
a TRADITION of INNOVATION



The advent of Tweetonomies?

■ Taxonomy

- hand-crafted hierarchical structure of concepts for classification

■ Folksonomy

- emerge when user collectively organize/classify resources
- conceptual structures and hierarchies on folksonomies (see e.g., [Schmitz, 2006], [Mika, 2007] and [Heymann, 2008])

■ Tweetonomy

- Do Tweetonomies emerge when users communicate and share information on SAS?
- To what extent does the type of stream aggregation and structure of stream aggregation influence emerging semantics?



Structure of SAS

- **Users, messages and content of messages**
- **Content of messages:**
 - ➔ words, URLs, and other user-defined syntax such as hashtags, slashtags or @replies.
- **Emerging collaboratively-defined syntax conventions make the structure of SAS more complex and dynamic than in other stream-based systems**



A network-theoretic model of SAS

- A **Social Awareness Stream** is a tuple

$$S = (U_{q1}, M_{q2}, R_{q3}, Y, ft)$$

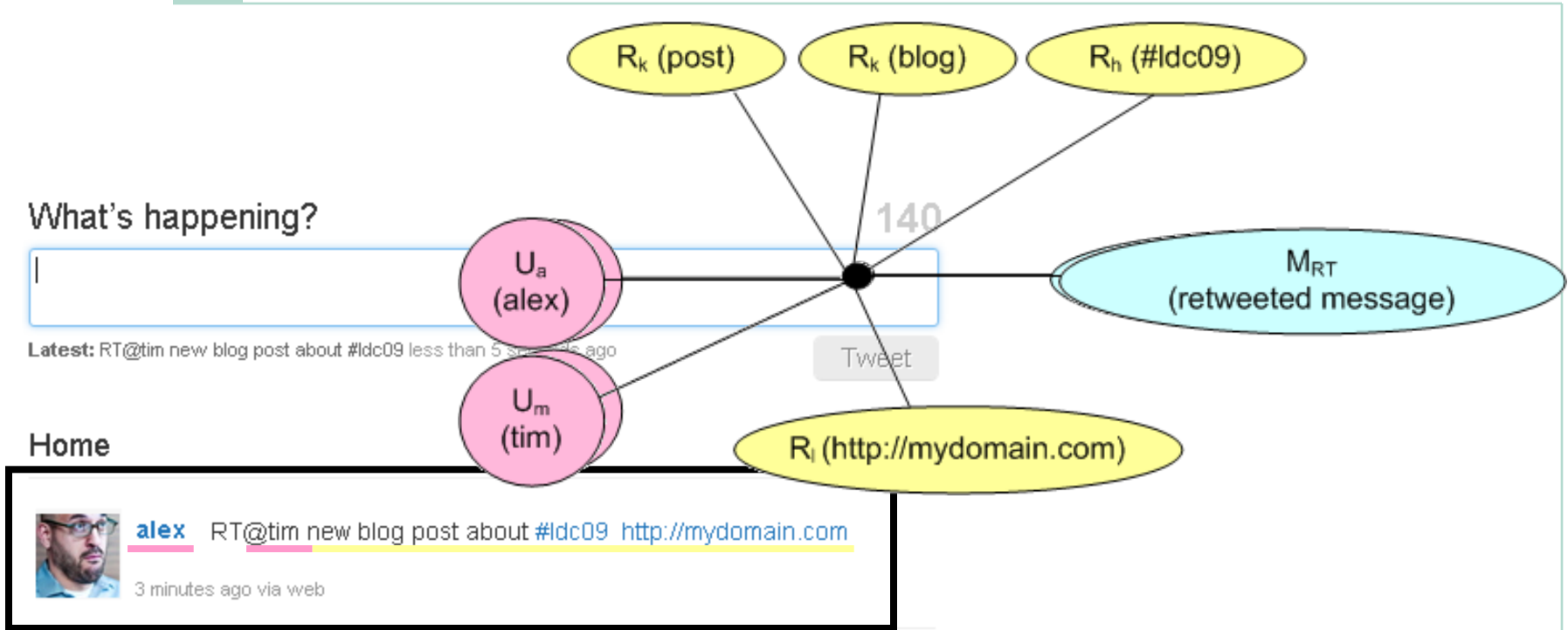
- U, M and R are finite sets whose elements are called users, messages and resources

- q1, q2, q3 are qualifiers

- Y is a ternary relation $Y \subseteq U_{q1} \times M_{q2} \times R_{q3}$

- ft is a function $ft : Y \rightarrow \mathbb{N}$

Example



SI_24Seven Woman suing Steelers' Holmes for alleged attack - A Florida woman is suing Pittsburgh Steeler Santonio... <http://bit.ly/dBTYHr>
23 minutes ago via API



Experiment

■ Aim

→ Explore nature of different stream aggregation types

■ Structure

→ Structural stream measures

■ Semantics

→ Simple network transformations



Dataset

■ 4 different stream aggregations from Twitter

■ Same topic

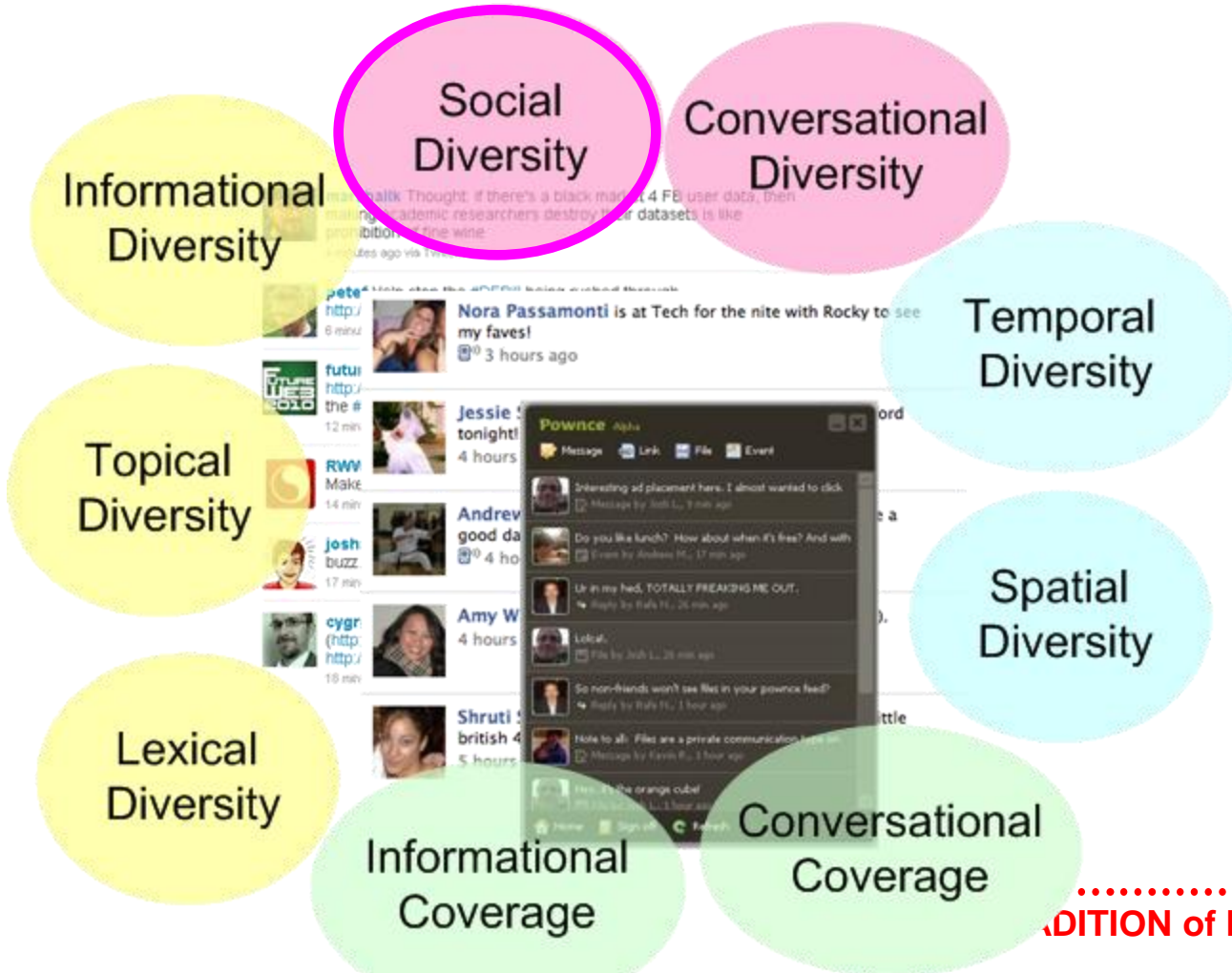
- Hashtag stream: #semanticweb
- Keyword stream: semanticweb and semweb
- User list stream: semweb user list from twitter user sclopit
- User directory stream: wefollow semanticweb directory

■ Same time interval

- 2 time intervals: 16th of Dec 2009 - 20th of Dec 2009 and 29th of Dec 2009 - 1st of Jan 2010



Structural Stream Measures (1)



.....
ADDITION of INNOVATION



Structural Stream Measures (2)

■ Social Diversity

→ How many different users participate in a stream?

→ Social variety:

$$SV_{pm} = \frac{|U_a|}{|M|}$$

$u \in U_a \dots$ authors
 $m \in M \dots$ messages

→ How balanced are their participations?

→ Social balance:

$$SB = - \sum_{u \in U_a} P(m|u) * \log(P(m|u))$$



Experiment

■ Aim

- Explore the nature of different stream aggregations

■ Structure

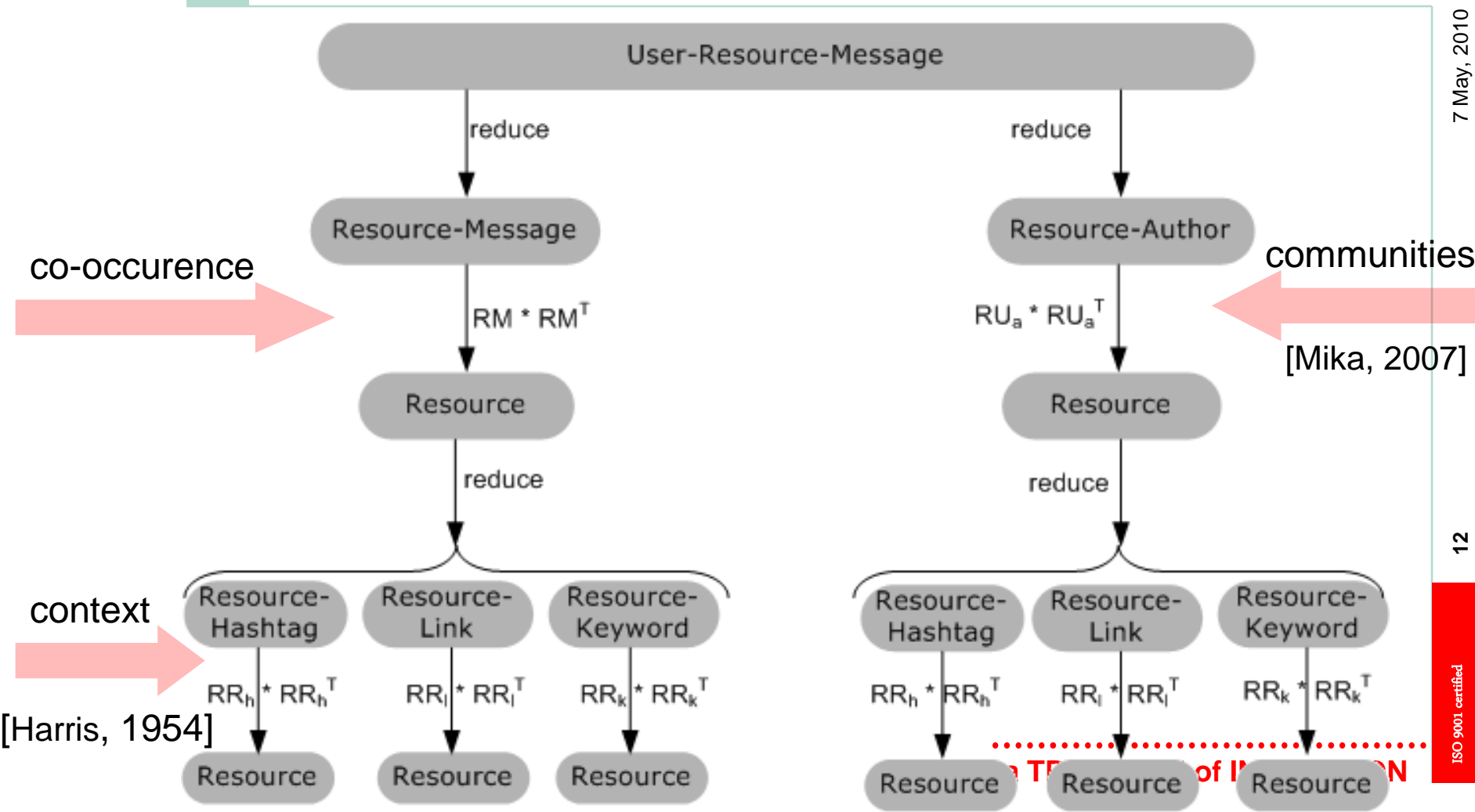
- Structural stream measures

■ Semantics

- Network-theoretic model of Social Awareness Streams
- 3-mode networks (users, resources and messages)
- Network transformations (projections) to obtain lower-order networks



Network Transformations





First Results (1)

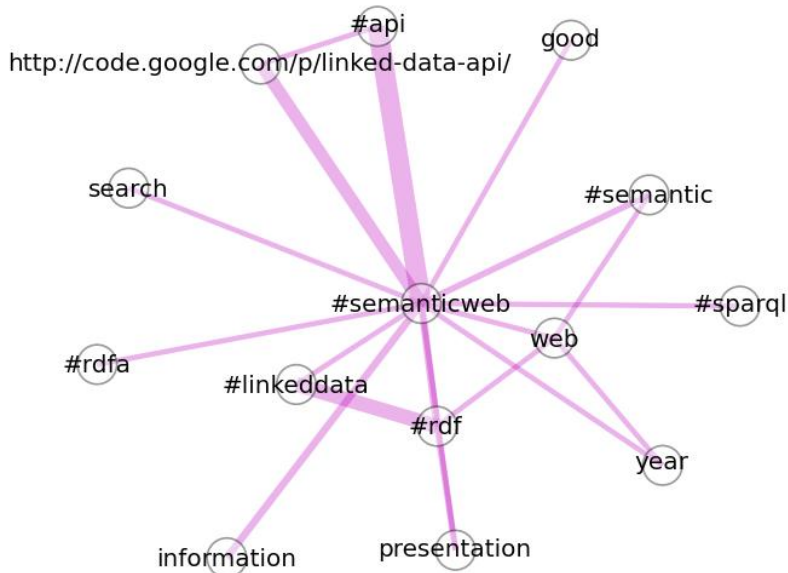
- **Type of stream aggregations influence stream structures**
 - ➔ Hashtags streams seem to be more informational than user list streams
 - ➔ Hashtag streams seem to be more social diverse than user list streams
 - ➔ User list streams seem to be slightly more conversational than hashtag streams



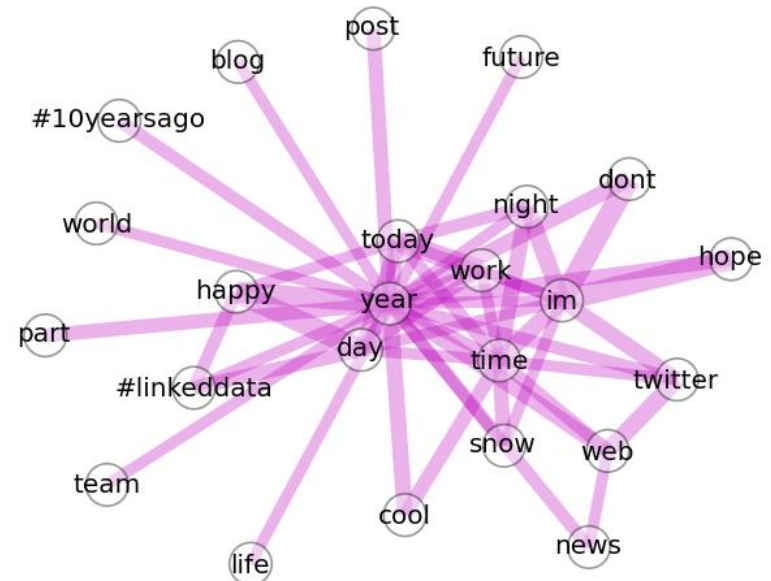
First Results (2)

■ Type of stream aggregations influence emerging semantics

➔ Hashtag stream aggregations are more robust against external disturbances than user list streams



Hashtag Stream
 $O_R(RU_a)S(R_h)$



User List Stream
 $O_R(RU_a)S(R_{UL})$

.....
a TRADITION of INNOVATION



First Results (3)

■ Type of network transformation influence emerging semantics

- Hashtags seem to be good context indicators
- Resource-hashtag networks reveal good latent conceptual structures



Conclusion

Theoretical Contribution

- **Network-theoretic model of SAS**
- **Structural Stream Measures**

Empirical Study

- **Do Tweetonomies emerge when users communicate and share information on SAS?**
 - ➔ Yes, latent conceptual structures can be observed
- **Does the type of stream aggregation and structure of stream aggregation influence emerging semantics?**
 - ➔ Yes, stream aggregation type influences structural properties and emerging semantics



References

- Z. Harris. Distributional structure. *The Structure of Language: Readings in the philosophy of language*, 10:146-162, 1954.
- P. Heymann, G. Koutrika, and H. Garcia-Molina. Can social bookmarking improve web search? In *WSDM '08: Proceedings of the international conference on Web search and web data mining*, pages 195-206, New York, NY, USA, 2008.
- P. Mika. Ontologies are us: A unified model of social networks and semantics. *Web Semant.*, 5(1):5-15, 2007.
- M. Naaman, J. Boase, and C.-H. Lai. Is it all about me? user content in social awareness streams. In *Proceedings of the ACM 2010 conference on Computer supported cooperative work*, 2010.
- P. Schmitz. Inducing ontology from Flickr tags. In *Proceedings of the Workshop on Collaborative Tagging at WWW2006*, Edinburgh, Scotland, May 2006.



Thank you!

claudia.wagner@joanneum.at

<http://clauwa.info/me>

<http://twitter.com/clauwa>

JOANNEUM



RESEARCH

.....
a TRADITION of INNOVATION

ISO 9001 certified