When Resources Collide: Towards a Theory of Coincidence in Information Spaces

Markus Luczak-Roesch
University of Southampton
Web and Internet Science
Southampton, United Kingdom
mlr1m12@ecs.soton.ac.uk

Ramine Tinati
University of Southampton
Web and Internet Science
Southampton, United Kingdom
r.tinati@ecs.soton.ac.uk

Nigel Shadbolt
University of Southampton
Web and Internet Science
Southampton, United Kingdom
nrs@ecs.soton.ac.uk

ABSTRACT

This paper is an attempt to lay out foundations for a general theory of coincidence in information spaces such as the World Wide Web, expanding on existing work on bursty structures in document streams and information cascades. We elaborate on the hypothesis that every resource that is published in an information space, enters a temporary interaction with another resource once a unique explicit or implicit reference between the two is found. This thought is motivated by Erwin Shroedingers notion of entanglement between quantum systems. We present a generic information cascade model that exploits only the temporal order of information sharing activities, combined with inherent properties of the shared information resources. The approach was applied to data from the world's largest online citizen science platform Zooniverse and we report about findings of this case study.

Categories and Subject Descriptors

H.1.1 [**Information Systems**]: Models and Principles—*Systems and Information Theory*

Keywords

Information cascades, social machines, information theory, socio-technical systems

1. INTRODUCTION

The World Wide Web is the largest openly accessible information space today. We can see an increasing number of examples where human individuals contribute to large-scale collective action by sharing information on the Web. This can be in case of a disastrous humanitarian event (e.g. the Haiti earthquake) or a political crisis (e.g. the Kenyan election) but also less critical situations deserving the spread of information of public interest (e.g. an actual traffic jam or cancelled train). These examples have in common that there is some higher-order event underlying the information sharing activity (e.g. coordinating help in disaster response or optimizing travel routes of people being affected by traffic disruptions). However, people are not necessarily talking with each other within a single Web-based application or triggered by explicit social links but about the same events instead using various information sharing services to broadcast information quickly (especially in critical situations when time to make decisions is rare). Motivated by this we argue that the problem solving capabilities in what is today known as Social Machines[26] is – rather than being an intentionally engineered piece of software – the substrate of accumulated human cross-system information sharing activities[17]. The formal definition, capturing, and analysis of those purposeful collective events emerging from coincidence, is the wider context of the work presented in this paper.

In 1935 Erwin Schroedinger stated that *"when two systems, of which we know the states by their respective representatives, enter into temporary [...] interaction due to known forces between them, [...], then they can no longer be described in the same way as before [...]", coining the notion of entanglement between quantum systems [24].* We pick this idea up for a new concept of coincidence as the foundation of collective higher order events in information spaces and raise the following hypothesis: **Every resource that is published in an information space, enters a temporary interaction with any other resource in that space once a unique explicit or implicit reference between the two is found.** System-specific and feature-rich methods for the analysis of the linking and flow of information in Web-based systems must be incomplete to describe this macroscopic state of the Web. What if we disregard all system-specific features and relationships and simply assume that any resource on the Web is potentially exposed to any other that is openly accessible? Does a form of entanglement exist between presumably independent streams of information sharing on the Web?

Summary of Contributions

This paper seeks to lay out foundations for a general theory of coincidence in information spaces, expanding on existing work on bursty structures in document streams and information cascades. As a first step we focus on temporary interactions between information that may allow us to investigate a form of entanglement in the future. We developed a generic information cascade model that exploits only the temporal order of content sharing activities that happen on the Web, combined with inherent properties of the shared resources. The model is configured by the stream of the time stamped resources (be it from live streamed data or replayed historic data), and the matching methods that are applied to analyse any inherent properties of the resources (e.g. the content). We used this approach to study the world's largest online citizen science platform Zooniverse. The system is an excellent example to study this organic information evolution: it does not support any typical social networking functionality; there is also evidence that no virtual social network emerged from people's interactions im-

Copyright is held by the International World Wide Web Conference Committee (IW3C2). IW3C2 reserves the right to provide a hyperlink to the author's site if the Material is used in electronic media.
WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2743973.

plicitly; content sharing contributes to cooperative hypothesis testing.

2. RELATED WORK

In response to the still popular initial characterisation of *Social Machines* as the "processes in which the people do the creative work and the machine does the administration" [5] a new line of research has been established [11, 20, 25] that most recently led to the following working definition: "Social machines are Web-based socio-technical systems in which the human and technological elements play the role of participant machinery with respect to the mechanistic realisation of system-level processes." [26] Our work refers to this definition and contributes insight into these "system-level processes" by expanding on existing research on information cascades in online communities.

The kick-off point for the extensive research on information diffusion online, which has been undertaken over the last decade, can be seen in the various approaches for the temporal analysis of Web crawls [7, 6, 9] and even more importantly Kleinberg's work on patterns of burstyness in document streams [13]. Bursts refer to periods of significantly high activity in continuous time stamped sequences of documents. They have become an accepted indicator for the appearance of a topic and can be used to infer meaning by analyzing the content in documents that belong to a particular burst. This method has significantly impacted how we study temporal properties of human-generated digital content networks [14, 22] but is also complemented by a much wider view on the role of bursts of activity in human behaviour [4].

Research on information cascades grew out of this work as an established means to study the propagation of information on the Web. Information cascade models have been applied to represent and study the structure of the blogosphere, to analyse the viral spread of news online, and to measure influence in political campaigns to name just a few prominent application examples [2, 10, 1]. Cascades are typically modeled as dynamic networks [23]. This means that one or more undirected sub-networks represent structures of explicit relationships between entities (nodes in the network) along which information could possibly diffuse (e.g. blog sites interlinked by blog roll features or users forming a following or friendship graph). An actual diffusion process is represented as a time-stamped directed overlay network. Each edge in the overlay network is directed from the "infector" node to the "infectee" node as well as labelled with the time when the diffusion was evidenced on the side of the "infectee" and the identifier of the diffusing information. Evidence for an infection is inferred based on features of the sub-network. Cascades have a single initiator but they can collide and merge when identifiers from different cascades are used in one node [16, 15, 3, 8].

With our work we return to Kleinberg's original approach [13] but apply it to resources that are openly shared on the Web. We assume that there is a natural information transmission capability on the World Wide Web that is not necessarily conditioned by explicit social links. Our investigation expands the approach in [13] and is focused on the structural properties of branching and merging cascades that are derived when different pattern matching algorithms are applied to the same document sequences.

By contextualising our alternative model with implicit but still purposeful collaborative work, we also touch the field of collective intelligence, human computation, and social computing. Work in this area typically focuses on the intelligence and problem solving capability that results from virtually organized groups working together towards a particular outcome and the coordination to optimise this [19, 27, 21, 12]. We instead want to expose the intelligence that lies in information on the Web that is linked because of coincidence rather than pre-configured conditionality or necessary a priori planning.

3. INFORMATION CASCADES: A TRANSCENDENTAL MODEL

As noted, we explore the possibility of abstracting the social context away from the technological substrate to understand the Web's intrinsic information cascades, considering not only local understanding of its use but also an abstract global view. This lets us propose a new model that we call *transcendental information cascades*. Informed by Kleinberg's work on burst structures in streams [13] it regards time as the only ascertainable condition for relationships between any two resources. Beside that we focus on coincidence of information sharing activities rather than socially-determined conditionality.

We define a *transcendental cascade* as a directed network. In contrast to the common information cascade model [8] we do not presume any sub-network to exist but only a set of *resources* (e.g. individual blog posts, microposts, forum entries, or Web pages). Resources are stamped with the time when they were shared. Nodes in the network are those resources from this set that contain one or multiple *cascade identifiers*. A cascade identifier is any unique pattern that is recognized by applying a *matching function* to the content or any other inherent properties of a resource (e.g. simple string matching algorithms to identify keywords in content). An edge exists between any two nodes that share a unique subset of the cascade identifiers that were found for them. This subset and any of its enclosed subsets must not be part of the identifiers featured by any node that was created in the time period between the two linked nodes were created. A node that contains a cascade identifier that was not detected for any other nodes before is called a *root* for this identifier. A node that has no outgoing edges is called a *stub*.

Formally a transcendental cascade is a tupel TC that comprises a set of nodes V, a set of edges E, a set of resources R and a set of matching functions F.

$$TC = (V, E, R, F)$$

The *resources* R in this case are defined by a unique identifier, the time when they were shared, and their content.

$$R = \{r_1, r_2, ..., r_m\}$$
$$r_i = (u_i, t_i, c_i),\ m, i \in \mathbb{N}, i \leq m$$

Complementary to resources exists the set of pre-configured matching functions F.

$$F = \{f_1, f_2, ..., f_n\}, n \in \mathbb{N}$$

A matching function can be any algorithm that is suited to analyze inherent properties of contents c_i of resources $r_i \in R$ for recognition and extraction of patterns. As mentioned before, this can be any text pattern such as keywords or phrases, but also patterns in images and videos (if the content is of that kind) and even more complex semantics or sentiments. We define a matching function $f_k \in F, k \in \mathbb{N}, k \leq n$ as:

$$f_k(c_i) = \begin{cases} \{i_1, i_2, ..., i_x\} & \text{if } f_k \text{ matches patterns} \\ & \{i_1, i_2, ..., i_x\} \text{ in } c_i \\ & x \in \mathbb{N} \\ \\ \emptyset & \text{otherwise} \end{cases}$$

By applying every matching function to all contents of resources $r_i \in R$, we derive a set of nodes V with one corresponding node for each resource with a non-empty set of cascade identifiers I_i. Each node $v_y \in V$ then is described by a unique identifier u_y, a time stamp t_y that marks when the node was published, and a set of cascade identifiers I_y. And each set of cascade identifiers I_i is given by the concatenation of the results of all matching functions in F applied to c_i.

$$V = \{v_1, v_2, ..., v_p\}$$

$$v_y = (u_y, t_y, I_y),$$

$$p, y \in \mathbb{N}, y \leq p$$

$$I_i = \{i_1, i_2, ..., i_o\}$$

$$= f_1(c_i) \cap f_2(c_i) \cap ... \cap f_n(c_i)$$

$$o = \sum_{x=0}^{n} |f_x(c_i)|$$

$$\Rightarrow \forall i_j \in I_i \exists f_k(c_i) \rightarrow i_j \in f_k(c_i), j \leq o$$

Edges e_z are directed from the source node with the identifier u_a to a target node with the identifier u_b. They exist between any two nodes $v_a, v_b \in V$ that have a common identifier subset Λ_z (link identifiers) within their respective sets of cascade identifiers I_a and I_b from which no identifier has been used by any other node with a time stamp between t_a and t_b.

$$E = \{e_1, e_2, ..., e_q\}$$

$$e_z = (u_a, u_b, \Lambda_z)$$

$$q, z \in \mathbb{N}, z \leq q$$

$$\Lambda_z = \{i_r |$$

$$i_r \in I_a \wedge i_r \in I_b,$$

$$\forall i_r \rightarrow V' =$$

$$\{v_c | v_c = (u_c, t_c, I_c), i_r \in I_c \wedge t_a \leq t_c \leq t_b\} = \emptyset,$$

$$v_c \in V, r \in \mathbb{N}, r \leq |I_b|\}$$

This cascade model yields different outputs depending on the data to hand (e.g. determined by the extent of the Web crawl), and the matching algorithms determining which cascade identifiers will be spotted (e.g. reuse of hashtags, URIs, quotes, images, or maybe exploiting wider semantics or sentiment). We call the activity when the next resource is analysed for encapsulated cascade identifiers and potential linking with a node in an existing cascade a *collision*. Successful collisions lead to the generation of new edges. A generic architecture, that shows how the data (either streamed or historic) and a set of matching algorithms form an input configuration to our approach, is depicted in Figure 1.

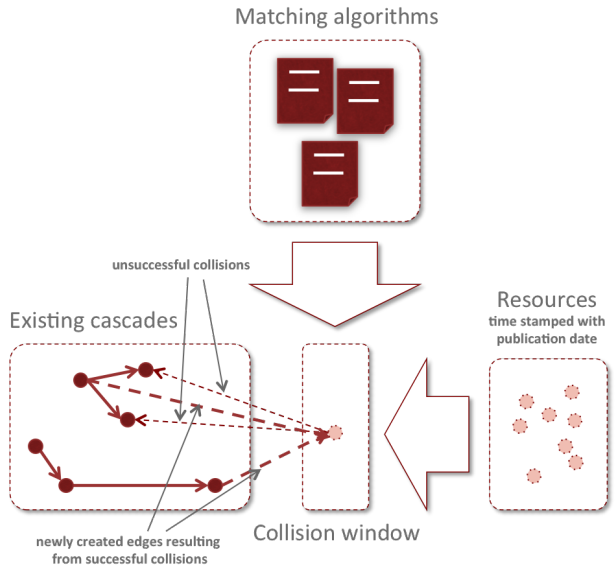

Figure 1: Generic resource collider architecture to generate transcendental information cascades.

A fictive example of a transcendental cascade based on our model is shown in Figure 2. Consider a system that features hashtags as an established form of identifying content patterns. The visualisation uses the following approach to represent distinct identifiers and time: Nodes are chronologically ordered alongside the horizontal dimension from left (the oldest node) to right (the most recent node); additionally nodes are ordered alongside the vertical dimension depending on the set of identifiers present in a node (each unique set is assigned to a distinct level). Consequently, the visualisation represents the content creation sequence *("#A") - ("#A#B") - ("#A") - ("#A") - ("#A#B#C") - ("#C") - ("#A") - ("#B#D") - ("#A")*.

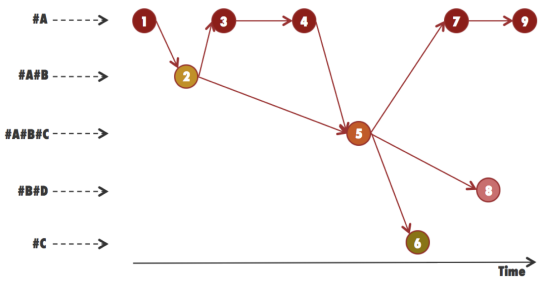

Figure 2: Example of a cascade that emerges along five different identifiers. #A, #B, #A#B#C, #B#D and #C are fictive hashtags (or hashtag combinations resepectively) treated as the indentifying content patterns

4. A CITIZEN SCIENCE CASE STUDY

Zooniverse is the world's largest citizen science platform today, in which volunteers complete tasks in aid of advancing the current state of scientific knowledge. Participants can sign up to a range of projects, which range form classifying the shape of a galaxy, to spotting animals in the Serengeti. There are over 30 different Zooniverse projects which share the same platform, and over 1 million active participants as of December 2014. In addition to the task

functionality, the Zooniverse platform also developed *Talk*, a forum system which supports regular forum posts, unlimited in length, and microposts, which are limited to 140 characters. Each project has it's own Talk system, but users are easily able to post across projects as the Zooniverse platform supports single sign-on.

As shown in [18], microposts dominate talk in all projects, accounting for more than 90% of all content elements. Moreover, the discussion patterns often do not resemble that of typical online communities; instead of *talking to each other*, in particular projects such as Planet Hunters, participants are just *talking about the same things* in a highly specific way. An example post from the Planet Hunters project hosted on the Zooniverse platform is shown in Figure 3.

Figure 3: Excerpt of an example thread on the Planet Hunters forums, showing the use of very specific terminology by participants to hypothesise about objects of interest.

With this initial study we seek to gain insight into basic structural and informational properties of the cascades created by our transcendental model. We applied the approach to two datasets representing chronologies of content sharing on the Zooniverse citizen science platform between 15/12/2010 and 16/07/2013: (1) a complete dataset of all forum and micropost entries within all projects on the platform consisting of 660,441 posts that have been contributed by 47,888 participants; (2) a complete dataset of all forum and micropost entries only within the Planet Hunters project comprising of 427,917 posts contributed by 32,805 participants.

For dataset 1, only the hashtag identifier was used as the matching function (A1) to identify cascades. *Hashtags* are supported in all projects hosted on the Zooniverse platform in the way as it is very common today for micropost platforms such as Twitter. This means that whenever a "#" is directly prepended to a character sequence without white space, a link to a list of all posts using this identifier is created. Participants in all projects use hashtags as a way to talk about specific features of the subject processed in the crowdsourcing task. To account for the impact different matching functions have on the resulting cascade in our model, dataset 2 was processed by applying two additional pattern matchers. The second pattern, A2, is used to match content that refers to specific object identifiers related to the images shown in Planet Hunters. Over the course of the Planet Hunters project participants started to link posts to related objects they were presented with during task completion by mentioning their identifiers. These identifiers feature a consistent pattern of the form "APH[0-9]*", so we developed a second string matching algorithm (A2) to use these patterns as cascade identifiers. Finally, the third pattern, A3 is related to another type of identifier used by the Planet Hunter community to refer to objects in specialised external datasets. The community have built up enough domain knowledge to hypothesise which *exoplanet* candidate (aka Kepler candidate) a particular light curve might belong to by mentioning the respective Kepler candidate ID ("KID[0-9]*"). Analogous to the matching of internal object identifiers we implemented a string matching algorithm for KID patterns (A3).

Basic properties of transcendental cascades

Table 1 lists basic structural properties for the hashtag cascades generated for dataset 1. We calculated the average links-per-node as well as the number of individual cascade roots and stubs. Most noteworthy the number of roots is by more than factor 4 higher than the number of stubs. That means that over time there must be significant amount of merges that lead to information, that is encapsulated in a particular hashtag, being absorbed (and lost) or combined with the information of other hashtags. This is also reflected by the proportion of 1.26 links per node.

Nodes	165,562
Links	208,122
Cascades	871
Roots	10,230
Stubs	2,200
Avg. links per node	1.26

Table 1: Basic structural cascade properties – complete Zooniverse dataset (dataset 1).

Table 2 lists the same properties for dataset 2 when the matching functions A1, A2, and A3 are applied individually and in combination (A1+A2+A3). In hashtags cascades (A1), each node contains 1.15 links. In comparison to this, APH (A2) and KID (A3) nodes feature only 0.56 and 0.65 links respectively. Being dominated by the amount of nodes in hashtag cascades, the combination of matching functions (A1+A2+A3) results in an average of 1.08 links per node. Computing the roots, we found that for hashtag cascades (A1), the amount of roots and stubs is much lower than in case of the APH or KID datasets, with APH featuring the lowest proportion of stubs per root (0.21). While this proportion is quite similar for hashtags (0.28), KIDs stand out in this case with a relatively high proportion of stubs per root (0.40).

	A1+A2+A3	**A1**	**A2**	**A3**
Nodes	89,023	66,616	11,032	16,728
Links	96,507	76,776	7,142	9,442
Cascades	2,784	185	1,519	3,291
Roots	25,560	1,458	14,124	10,127
Stubs	6,350	416	2,931	4,097
Avg. links per node	1.08	1.15	0.56	0.65

Table 2: Basic structural properties when different matching functions and matching function combinations are applied to create transcendental cascades from dataset 2.

Retrieving and visualising transcendental cascades

We developed an intuitive search engine prototype that allows to

perform simple keyword search for retrieving relevant coincidental information sharing events. For the visual representation we adapted the D3 Sankey diagram[1], which allows us to demonstrate how cascades merge and branch. In Figure 4 we show a cascade retrieved from the full Zooniverse hashtag cascade database (dataset 1). The cascade groups content elements from the Notes from Nature Zooniverse project[2] with all tags (node labels) being related to specific forms of flying insect traps. This shows that our transcendental cascade model can be used as an indexing mechanism to retrieve content related to a particular topic, which was distributed across a larger system without any explicit relationships.

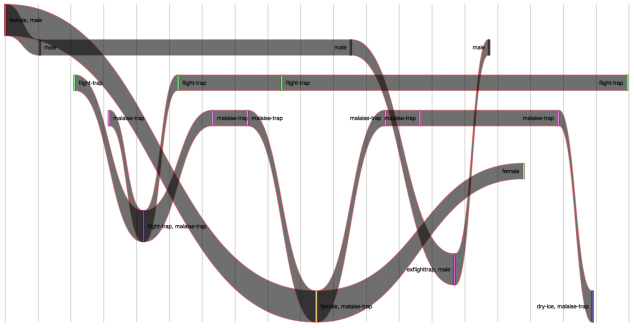

Figure 4: Information cascade linking content about flying insect traps in the Notes from Nature project discussion forums on the Zooniverse platform.

5. DISCUSSION

In this section we examine the results of our Zooniverse cascade case study and consider how our transcendental model for constructing cascades offers an alternative way to understand information flow on the Web.

Structural properties of transcendental cascades

Informed by the varying amount of roots and stubs as well as the different proportion of links per node for different matching function configurations, we believe that there is value in not only studying the sub-structures that emerge from information cascades, but even more by exploring the repetitive patterns of sub-structures. This differs from existing research [13]. Rather than examining the time difference between individual content elements, studying burstiness with respect to repetitive sub-structures seems to be promising to study phenomena like outperforming, drift or oppression of particular topics over time. This can become an important means to understand the evolution of online campaigns and virtual mass coordination from a macroscopic viewpoint and independent from preconceived social determinants.

Informational properties of transcendental Cascades

Our study raises questions about information gain and loss, and if information cascades are a way to observe and measure this. We found that there are varying distributions of roots and stubs in cascades. In some cases, we see more roots than stubs, which suggests that information that goes into cascades as distinct input, does not come out. This means there is information loss or information gets absorbed when a particular hashtag wins over others on a particular topic (convergence) for example.

We suggest that an information cascade can be considered as an entity that flows through the Web, channelling and preserving information across time. It therefore has storage and transfer capacity, and as a result is an important aid particularly for distributed communities with few communally-created information storage facilities capable of allowing access to information in a timely manner at the point at which it is needed. Some, but not all, input signals become output signals, so a body of information can evolve over time. Information loss may correspond to information ceasing to be current, or alternatively a cascade might branch to create divergent cascades whose combined capacity may make up for apparent local losses.

Scaling cascades up to the entire Web and prospects for new search engines

Whilst the analysis performed in this paper focused on preselected patterns on a well-structured, known dataset, we seek to consider how information cascades can be observed via an organic process, rather than via a predefined model or set of criteria. We work towards the construction of transcendental cascades at Web-scale, independent of system borders and entity types. We envision this as an alternative – temporal – way to consider the Web graph, not as a static network of hyperlinks, but as a network of entities flowing and interacting – the Web as a stream. The potential is to provide a new paradigm for search engines, to lead information consumers to a particular sequence in the information flow, rather than returning a ranked list of documents.

Yet, such a vision raises several challenges. Computationally, the construction of cascades, in real-time or as an offline batch job is challenging. It would require a large amount of computational resources, especially if no predefined pattern is to be used as the criteria. As Web crawling research has shown [7, 6, 9], not all resources that are published are timestamped in a detailed way. It might rather be that the most detailed temporal information is given by the date when a comprehensive Web crawl was made and new resources were found for the first time. As a consequence, a large number of resources would have the same timestamp, thus requiring heuristics to determine which one is the first to be analysed for collision with resources in existing cascades. Additionally, we have to investigate strategies for dealing with disappearing resources that were present in old crawls but were deleted so more recent crawls will not contain them anymore.

6. CONCLUSIONS

In this paper we took a first step towards an alternative theory for information diffusion online. Rather than proving conditionality with feature-rich probabilistic methods, we evaluated the suitability using inherent properties of shared content. A cascade is thus the results of content that is set out in an information space, in our case the World Wide Web, and the technical capabilities at hand to analyse that content for patterns of similarity. We formalised a generic model that accounts for all these aforementioned aspects. This includes not only the model of the directed network that represents a cascade, but also the function that describes the transformation of resource properties into cascade identifiers in an abstract way.

We contextualised our work in the beginning with the ambitious question whether the Web features an intrinsic problem solving potential so that the overall information propagation behaviour forms giant purposeful events. The analysis we undertook so far is only a first step towards a satisfying answer to this question, but it sets the line for future work and significantly informs the experimental design in this space.

Future Work

The findings in this paper suggest two promising directions for immediate next steps: First, it is necessary develop scalable methods for the analysis of cascade motifs. This needs to go up to the extend to identify the optimal segmentation of cascades into all those mo-

[1] `http://bost.ocks.org/mike/sankey/`
[2] `http://notesfromnature.org`

tifs that consist of a maximum number of nodes without containing any repetitive non-trivial motifs (non-trivial motifs are sub-graphs with > 1 nodes).

Second, we are going to quantify the informational properties of the cascades by measuring identifier entropy and information capacity. This will allow us to be more precise about the role of specific node types and cascade motifs that indicate information gain and loss.

7. ACKNOWLEDGEMENTS

This work was supported by the EPSRC Theory and Practice of Social Machines Programme Grant, EP/J017728/1. Markus Luczak-Roesch wants to thank Martin Heanue for conversations that inspired thoughts, which found their way into this paper.

8. REFERENCES

[1] Adar, E., and Adamic, L. Tracking information epidemics in blogspace. In *Web Intelligence, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference on* (Sept 2005), 207–214.

[2] Adar, E., Zhang, L., Adamic, L. A., and Lukose, R. M. Implicit structure and the dynamics of blogspace. In *Workshop on the weblogging ecosystem*, vol. 13 (2004), 16989–16995.

[3] Bakshy, E., Hofman, J. M., Mason, W. A., and Watts, D. J. Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining*, ACM (2011), 65–74.

[4] Barabasi, A.-L. The origin of bursts and heavy tails in human dynamics. *Nature 435*, 7039 (2005), 207–211.

[5] Berners-Lee, T., and Fischetti, M. *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. HarperInformation, 2000.

[6] Bharat, K., Chang, B.-W., Henzinger, M., and Ruhl, M. Who links to whom: Mining linkage between web sites. In *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, IEEE (2001), 51–58.

[7] Brewington, B. E., and Cybenko, G. Keeping up with the changing web. *Computer 33*, 5 (2000), 52–58.

[8] Cheng, J., Adamic, L., Dow, P. A., Kleinberg, J. M., and Leskovec, J. Can cascades be predicted? In *Proceedings of the 23rd international conference on World wide web*, International World Wide Web Conferences Steering Committee (2014), 925–936.

[9] Fetterly, D., Manasse, M., Najork, M., and Wiener, J. A large-scale study of the evolution of web pages. In *Proceedings of the 12th international conference on World Wide Web*, ACM (2003), 669–678.

[10] Gruhl, D., et al. Information diffusion through blogspace. In *Proceedings of the 13th International Conference on World Wide Web*, WWW '04, ACM (New York, NY, USA, 2004), 491–501.

[11] Hendler, J., and Berners-Lee, T. From the semantic web to social machines: A research challenge for ai on the world wide web. *Artificial Intelligence 174*, 2 (2010), 156–161.

[12] Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., and Horton, J. The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work*, ACM (2013), 1301–1318.

[13] Kleinberg, J. Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery 7*, 4 (2003), 373–397.

[14] Kumar, R., Novak, J., Raghavan, P., and Tomkins, A. On the bursty evolution of blogspace. *World Wide Web 8*, 2 (2005), 159–178.

[15] Leskovec, J., Backstrom, L., and Kleinberg, J. Meme-tracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, ACM (New York, NY, USA, 2009), 497–506.

[16] Leskovec, J., McGlohon, M., Faloutsos, C., Glance, N. S., and Hurst, M. Patterns of cascading behavior in large blog graphs. In *SDM*, vol. 7, SIAM (2007), 551–556.

[17] Luczak-Rösch, M., Tinati, R., O'Hara, K., and Shadbolt, N. Socio-technical computation. In *CSCW'15 Companion*, ACM (2015).

[18] Luczak-Rösch, M., Tinati, R., Simperl, E., Kleek, M. V., Shadbolt, N., and Simpson, R. Why won't aliens talk to us? Content and community dynamics in online citizen science. In *Eighth International AAAI Conference on Weblogs and Social Media* (2014).

[19] Malone, T. W., Laubacher, R., and Dellarocas, C. Harnessing crowds: Mapping the genome of collective intelligence.

[20] Meira, S. R., Buregio, V. A., Nascimento, L. M., Figueiredo, E., Neto, M., Encarnacao, B., and Garcia, V. C. The emerging web of social machines. In *Computer Software and Applications Conference (COMPSAC), 2011 IEEE 35th Annual*, IEEE (2011), 26–27.

[21] Minder, P., and Bernstein, A. Crowdlang-first steps towards programmable human computers for general computation. In *Human Computation* (2011).

[22] Myers, S. A., and Leskovec, J. The bursty dynamics of the twitter information network. In *Proceedings of the 23rd international conference on World wide web*, International World Wide Web Conferences Steering Committee (2014), 913–924.

[23] Qu, Q., Liu, S., Jensen, C. S., Zhu, F., and Faloutsos, C. Interestingness-driven diffusion process summarization in dynamic networks. In *Machine Learning and Knowledge Discovery in Databases*. Springer, 2014, 597–613.

[24] Schrödinger, E. Discussion of probability relations between separated systems. In *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 31, Cambridge Univ Press (1935), 555–563.

[25] Shadbolt, N. R., et al. Towards a classification framework for social machines. In *Proceedings of the 22nd international conference on World Wide Web companion*, International World Wide Web Conferences Steering Committee (2013), 905–912.

[26] Smart, P. R., Simperl, E., and Shadbolt, N. A taxonomic framework for social machines. In *Social Collective Intelligence: Combining the Powers of Humans and Machines to Build a Smarter Society*, D. Miorandi, V. Maltese, M. Rovatsos, A. Nijholt, and J. Stewart, Eds. Springer, Berlin, Germany, 2014, 51–85.

[27] Von Ahn, L. Human computation. In *Design Automation Conference, 2009. DAC'09. 46th ACM/IEEE*, IEEE (2009), 418–419.