Noname manuscript No.
(will be inserted by the editor)

What an Entangled Web We Weave: An Information-Centric Approach to Time-Evolving Socio-Technical Systems

Markus Luczak-Roesch · Kieron O'Hara · Jesse David Dinneen · Ramine Tinati

the date of receipt and acceptance should be inserted later

Abstract A new layer of complexity, constituted of networks of information token recurrence, has been identified in socio-technical systems such as the Wikipedia online community and the Zooniverse citizen science platform. The identification of this complexity reveals that our current understanding of the actual structure of those systems, and consequently the structure of the entire World Wide Web, is incomplete. Here we establish the principled foundations and practical advantages of analyzing information diffusion within and across Web systems with Transcendental Information Cascades, and outline resulting directions for future study in the area of socio-technical systems. We also suggest that Transcendental Information Cascades may be applicable to any kind of time-evolving system that can be observed using digital technologies, and that the structures found in such systems consist of properties common to all naturally occurring complex systems.

Keywords information · philosophy · temporal data mining · socio-technical systems · information theory · information systems · social network analysis · complexity science · network science

---

M. Luczak-Roesch
Victoria University of Wellington
School of Information Management
E-mail: markus.luczak-roesch@vuw.ac.nz *Present address:* of F. Author

K. O'Hara
University of Southampton
Web and Internet Science Group
E-mail: kmo@ecs.soton.ac.uk

J. D. Dinneen
Victoria University of Wellington
School of Information Management
E-mail: jesse.dinneen@vuw.ac.nz

R. Tinati
Microsoft Corp
E-mail: ratinati@microsoft.com

1 Introduction

The World Wide Web (for short the Web) is the largest socio-technical system in existence, a system in which very large numbers of social agents and technical components act and interact. Despite the Web's apparent randomness and unpredictability, graphing its hyperlinks allows the detection of some universal properties that give a relatively precise answer to the question of what its overall structure may be (Barabási et al., 2000; Adamic and Huberman, 2000; Broder et al., 2000). Today, more people than ever before are able to publish content on the Web: from short posts on microblogging platforms to content in online communities, blog and Wiki articles, and HTML Web pages. This growth has been supported in part by an abundance of tools that make it easier for people to create and publish content without high levels of expertise and training. Use of such tools usually results in content that is explicitly time-stamped, which has increased the salience of temporal relationships for understanding the relations between pieces of content. Further relationships still reside in patterns *within* the shared content (e.g., categories, tags or hashtags, mentions of usernames). However, in contrast to traditional hyperlinks, those relationships are usually only explicit within one particular system; for example, hashtags within a microblogging platform are only for internal uses on a platform and do not link to the content on other platforms, turning the Web into a series of walled gardens.

In this article we argue on principled and practical grounds that increased temporal dynamics and implicit content relationships are challenging our understanding of the actual structure of the Web in particular and potentially of all time-evolving socio-technical systems, and that a new method for modeling that structure is needed to overcome these challenges. Our argumentation employs a dialectic of literature on the philosophy of truth and science as well as analytical methods for the study of information diffusion, Web graphs and social networks in order to make a more general case for changing the current view to the actions of human collectives in the digital. We present a method for modeling information diffusion by constructing Transcendental Information Cascades (TICs), which breaks with the *causality assumption* implicit in many of todays analytical methods and thus allows the capture of novel dimensions of complexity of information-sharing from a macroscopic perspective. We also discuss the theoretical contributions of modeling information in this way.

Our work adds to prior research emphasizing the role of activity and interaction sequences as the fundamental unit of analysis in socio-technical systems (Keegan et al., 2016), but further seeks to widen the scientific discourse on that matter. We acknowledge the value that lies in prior analytical methods mining causal structures and behavioral motifs from sequential data, but we also highlight that there is benefit in understanding the layers of complexity present among *low-level coincidences* in the various information sequences produced by socio-technical systems. What we describe here is directly related to research on socio-technical systems as investigated in computer-supported

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2789v2 | CC BY 4.0 Open Access | rec: 15 Apr 2018, publ: 15 Apr 2018

collaborative work (CSCW), Web Science, and computational social science (CSS). However, the methods we have developed and the observations we have made may also have wider implications for the study of any kind of system in which information is sequentially ordered. Examples include physiological time series such as EEG and ECG, historical text archives, and literary traditions.

The remainder of this article is structured as follows. In Section 2 we discuss why context-rich data analysis methods are limited in describing the organic nature of information in modern socio-technical systems because of their reliance on single channels of conversation, particular binary relationships, and causality between individuals' activities. In doing this we establish the notion of a complex artifact defined by relationships of meaning, which can be captured by a novel method for modeling information in *Transcendental Information Cascades* (TICs). After introducing TICs we describe several of their properties that can be leveraged to investigate information dynamics in socio-technical systems in a novel way, avoiding the problems context-rich methods encounter when the aim is to get a complete picture of the macroscopic state of an entire system such as the Web. In Section 5 we describe two cases where we applied TICs to data sets obtained from the world's largest online citizen science platform Zooniverse[1] and the English version of the Wikipedia[2] online encyclopedia. We then discuss the findings from those two cases in relation to existing literature on information clustering, temporal data mining, computer-supported cooperative work, and theories of social and information systems. In Section 6 we discuss the theoretical contributions of TICs, noting four reasons for thinking they may provide desirable theoretical underpinnings for the study of socio-technical systems. We conclude the article by summarizing the contributions we sought to make and outlining a few directions for a research agenda for Transcendental Information Cascades.

2 When socially determined network models fall short

Since the inception of the World Wide Web 29 years ago, information-sharing behavior has evolved from a relatively static picture (a rather small number of content contributors authoring individual hypertext pages with embedded hyperlinks) to one where a) content is shared at very high (and still growing) rate and b) links between content elements are increasingly implicit, as they emerge from common metadata (e.g., categories) or patterns in the actual content such as hashtags or username mentions. Many of the analytical methods that are applied to socio-technical systems (i.e., those found on the World Wide Web and most modern business information systems) are based on the assumption that there must be some retrievable snapshot of the structure underlying the behavior of the user base (e.g., a human collective using the Web) that can be used to infer causality (e.g., a social network graph). We call this assumption the *causality assumption* and argue that while it may hold for many particular

---

[1] https://www.zooniverse.org

[2] https://en.wikipedia.org

systems, it is unlikely to hold across the heterogeneous systems that comprise the World Wide Web for example, or even all those systems in action within a single organization. Thus, the assumption poses an obstacle to macroscopic views of information and information structures.

2.1 The problem by example

To demonstrate the limitations of the causality assumption, let us consider the example of digitally-mediated disaster response. The earthquakes in Haiti (2010) and Nepal (2015), the political crises in Congo and Somalia, and the recent Ebola outbreak in West Africa (2014-16) are representative cases where dedicated Web applications were combined with general social media platforms to facilitate effective crisis response supported by the opportunistic gathering of crisis-related information. The information relevant to a particular crisis goes well beyond that shared by individuals to support crisis management directly; depending on the purpose of the content creator, it can be intentional (e.g., contributions via an instance of the Ushahidi tool suite) or accidental that information is relevant to relief coordination (e.g., a micropost on Twitter about one or more cancelled flights from or to a particular airport could unintentionally provide relevant information for aid workers who have to reach a crisis region from outside of that area). In such cases we see that a) crisis-related information does not necessarily reside on a single platform and b) even when selecting a particular platform to capture crisis related information, the relevance depends on the *content* rather than the social networks that platform is supporting. As depicted in Figure 1 collective action is more salient when viewing content dynamics over time rather than by explicit social structure of a single system (Lee and Paine, 2015). It is more likely that the relevant information about an event that affects many individuals - particularly heterogeneous individuals who have only weak links with the majority of others affected - goes beyond any individual communication channel.

The intuition behind our argument is that for a human conversation, or collection of conversations connected by a particular topic, the media of those conversations are of secondary importance to the participants (though of course, recalling McLuhans famous insight, this is not to say that the medium is irrelevant to understanding the message (McLuhan and Fiore, 1967)). Human interlocutors are swayed by many factors in their choice of medium for a particular communicative act, from opportunism, to habit, to the need for security and/or anonymity, to the devices in their (and their interlocutors) possession, to requirements for synchronous discussion, to the need to reflect and to gather information before communicating, to whether the interlocutors are communicating in official or private capacities, etc (Kraut et al., 2012). Someone who needs to communicate with someone else will use whatever is at hand, which may entail multiple venues rather than intentionally restricting the conversation to a single channel such as Twitter or email. A conversation made up of many communicative acts may therefore take place across many different me-

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2789v2 | CC BY 4.0 Open Access | rec: 15 Apr 2018, publ: 15 Apr 2018

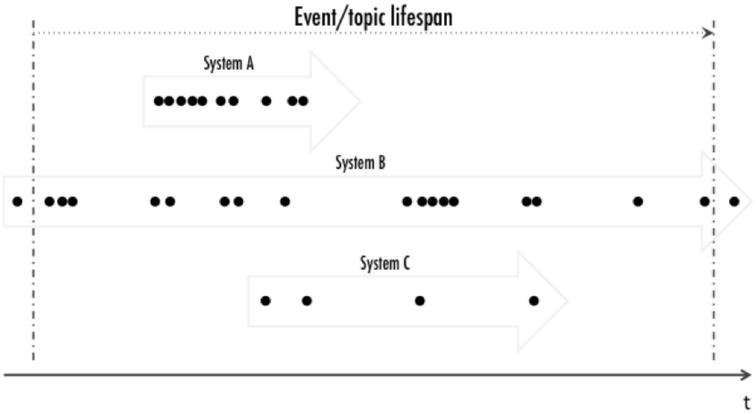

Fig. 1 Information naturally resides in an ecosystem and is emitted at varying frequencies. The accumulated information that is relevant for an event or topic forms the implicit collective action related to it.

dia, and this will be compounded when we aggregate conversations to try to set out an inclusive information picture. Many communicative acts will take place in face-to-face speech or in other unrecorded and unrecordable ways, and so these are not going to be captured. However, a method of recording that can encompass a range of media, not restricted to any individual channel such as a social networking site or a microblogging site, will, all things being equal, capture more of a conversation between diverse interlocutors. To take an obvious example, a tweet from President Trump might directly cause an editorial in the *New York Times*, and an official response on video from the North Korean leadership on a government website. Neither of these need necessarily be immediately detectable from analysis of Twitter.

Furthermore, by examining multiple media, groups of conversations will be connectable via the exogenous events that coincidentally motivate them. This is because the occurrence of a major event, such as a crisis (e.g., an earthquake) will trigger a number of independent conversations using similar vocabulary and identifiers on similar timescales. The aggregate of these conversations may be of great importance to crisis managers or rescue workers, but not visible when focusing on individual sources of data with particular models of discussion embedded into particular information infrastructures. What is of interest, then, is the wider collective discussion, a series of conversations connected only by the basic relationships of being about the same thing and taking place during a key time period. Connecting conversations on a single channel will be facilitated by the resources of that channel, of course, but a parallel argument shows that such a narrow focus will not only capture a fraction of each conversation at best, but will also cover only a fraction of the total number of relevant conversations across the Web.

The causality assumption makes single-channel analyses useful, of course; making contextual assumptions about particulars of a conversation, for example about the connection between two communicative acts (e.g., it may be inferred from the infrastructure of a channel that C is a reply to C), is easier if the data are taken from a single channel (e.g., only one channel's infrastructure must be described to make the inference). The opposite is true for multi-channel views, in which any two communicative acts that share a vocabulary and are closely connected in time cannot be inferred to be dialogue because they may be merely coincidental. The analyses of multi-channel models we propose constructing are therefore necessarily low in context, at least initially, and especially so when maximally inclusive, in which cases they may consist of both genuine conversations and coincidentally related communications. However, further analyses can re-introduce contextual factors to the initial low-context model, and can then be used to pare the structure down to something more intentional. For example, after constructing the model, known factors like friendship relationships can be re-introduced (e.g., by weighting edges in the network) to allow the inference of causality. This retains the benefits of high-context analyses, when they are possible, while avoiding problems of the causality assumption, and only in such cases can we analyse a topic or conversation across channels as we have argued for. Traditional methods in this space tend to assume that the available data, for example from a social networking site, capture the complete conversation and its context. This is a handy assumption for researchers, but it is limited in practice and unlikely to reflect the understanding of the participants.

2.2 The epistemological trap of system-specific analyses

At least three different principled problems underlying the practical ones outlined above have been described recently. First is a doubt about what real insight social network analyses and the study of social contagions (i.e., traditional analyses) can provide given that actors in those networks regularly overestimate the commonality of other actors' opinions or actions, for example because of misleading local maxima, which then influences their behaviour (Lerman et al., 2016). In other words, each actor's limited view of their network can cause them to produce network artifacts that serve as red herrings for and cause inaccuracies in traditional analyses of that network. Second, the simplified assumption that information spreads epidemically like viruses has created a wealth of analyses, including numerous current approaches to modeling information diffusion, that fail to account for other relevant factors, like the complex interplay of human cognitive limits and the algorithmic governance for content delivery imposed by platform providers (Lerman, 2016).

The first two problems discussed are caused by the properties and limitations of traditional analyses. The third problem, however, is bias in the study of such systems; namely, "our selective observation of successful [cases] provides us with a false narrative of [the] underlying causes" (Cebrian et al., 2016). To phrase it differently, analysts' definitions of measures of success (e.g., of successful information diffusion within a network) determine what they find in a network and thus the narrative they construct about causality within that network, which in turn skews the public perception of the digital sphere.

What we suggest is that these three principled problems cause an *epistemological trap*, wherein an analyst cannot know, for any particular informational pattern passing or being passed through a socio-technical system, whether it is any more than *suppositio materialis* (Tarski, 1944). In other words, because of the aforementioned human cognitive processes and factors motivating individual sharing activities excluded from such analyses, the shared meaning and significance of a term used by a collective of individuals (on Twitter, for example) will be impossible for the analyst to access. This inaccessibility entails that relationships manifested, for example, in a diffusion network, do not necessarily have any defining or designating reference to the knowledge object itself or some intensional semantics at a macroscopic scale (Whitehead and Russell, 1912), and thus are inscrutable. We conclude from these problems that traditional analytical methods relying on system-specific digital structures (e.g., social network analyses) are in principle precluded from knowing they have accurately represented the socio-technical systems they examine. While such methods function sufficiently well in some applied contexts (e.g., for predicting user behavior in online advertisement), their principled limitations entail that they do not, and cannot, satisfy traditional notions of truth (Popper et al., 1972), a necessary goal for scientific work. In other words, such analytical methods are epistemologically trapped.

One hopeful question might then be: is there an analytical method for capturing the structure and dynamics of a socio-technical system that is resilient against these aforementioned issues and therefore evades the resulting limitations?

2.3 Acknowledging complexity beyond pure reason: an entangled Web?

Given the above problems, our aim in developing such a desirable analytic method is to consider how we can understand evolving information in abstraction from both the social networks through which they flow and the system-specific digital traces of the social context. As outlined before, such consideration of socio-technical systems implies that system-specific data is incomplete to describe the macroscopic state of the space of all relevant information. Unconventional or hidden relationships between information, which would usually appear as noise relative to the explicit social relationships between the originators of the information, may be very influential despite, or independently of, the social structures created and curated by networking systems. Hence, we suggest separating the social context from the technological substrate to understand the Web's contribution *qua* abstract information space to the evolution of information. Whereas research on collective intelligence and human computation typically focuses on groups working explicitly or implicitly together towards a particular outcome and the coordination to optimize this (Malone et al., 2009; Woolley et al., 2010; Quinn and Bederson, 2011), in this research the goal is to expose the resources produced by accumulated human activities on the Web while minimizing the presuppositions about causality or the communities and systems (i.e., single channels) in which they take part.

Towards this goal, an analytical method has been proposed in (Luczak-Roesch et al., 2015b) and (Luczak-Roesch et al., 2015c). This method produces Transcendental Information Cascades, directed networks, formalised below, consisting of information co-occurrences characterized solely by time and inherent properties of pairs of content resources. The cascades are referred to as *transcendental* in Kant's sense of the word, namely, attempting to understand the conditions of knowledge itself (Kant, 1934). This means that not all such networks represent underlying purposeful activity; recurrence may simply be coincidence, but they do present a distinct set of properties of the macroscopic informational state of any socio-technical system like the Web. The relationships within and between such networks will reveal linking structures hidden from the system-specific point of view as well as temporal dependencies, describing a kind of *entanglement* between content resources.

3 The Transcendental Information Cascade formalism

We will now turn to the technicalities of Transcendental Information Cascades (Luczak-Roesch et al., 2015b,c) in order to demonstrate that there is already analytical capacity to capture the organic informational state of socially constructed information in systems, including those within other systems and across multiple channels. This shall demonstrate what tools or techniques may help understand this *entangled Web* independently of curated networks, and provide a reflection on two cases in which Transcendental Information Cascades are key to uncover otherwise hidden relationships between Web resources.

A *Transcendental Information Cascade* is defined as a directed network that is constructed from a sequence of *resources* (e.g., emails, blog posts, microposts, online forum entries, pages of a book, shared photos, time window snapshots of a continuous signal). Resources are ordered within the sequence (by the time of their occurrence from the oldest to most recent). Nodes in the network are those resources from this set that contain one or multiple *cascade identifiers*. A cascade identifier is any unique informational token that can be isolated by applying some *information extraction algorithm* to the original content of the elements of the sequence (e.g., natural language processing to identify unique words, phrases or entities in texts, or the color spectrum of specific areas in images). An edge exists between any two nodes that share a unique subset of all the cascade identifiers found in a sequence (e.g., two posts share some number of phrases, two images feature the same color spectrum in the same area), but only if no cascade identifier of this subset occurs in any element of the sequence between the two linked elements (i.e., assuming ordered elements A, B, and C, hashtags shared between A and C produce an A-C edge only when they are not also shared by B). Figure 2 shows a generalized example of this process, and examples with more detailed visualizations are provided below.

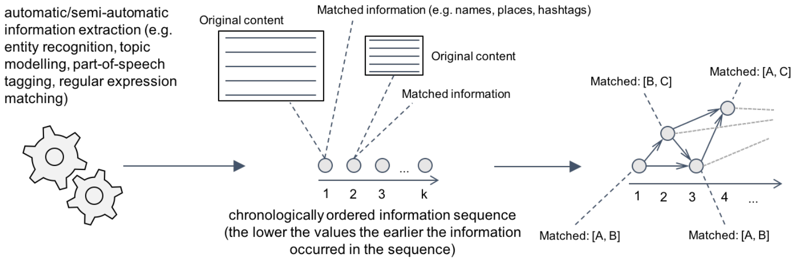

Fig. 2 Example of the method to construct Transcendental Information Cascades from an input sequence. For the sake of simplicity and illustrative purposes we assume that the output network was constructed by applying basic text pattern matching (in particular matching all capital letters) to textual content and that the text matched for the first four resources in the sequence was [‘A’,‘B’], [‘B’,‘C’], [‘A’,‘B’], and [‘A’,‘C’].

Formally a transcendental cascade is a tuple 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.

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

Resources R are defined by a unique identifier u_i, an ordering index t_i (e.g., the time when a resource was shared), and their content c_i.

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

Complementary to resources exists the set of matching functions F.

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

A matching function can be any algorithm that is suited to extract information from c_i of resources $r_i \in R$. We define a matching function $f_k \in F, k \in N, k \leq n$ as:

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

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, the ordering index t_y, and a set of cascade identifiers

I_y. And each set of cascade identifiers I_i is given by the union of the results of all matching functions in F applied to c_i.

$$\begin{aligned} V = \{v_1, v_2, ..., v_p\}, v_y = (u_y, t_y, I_y),\ p, y \in N, y \leq p \\ I_i = \{i_1, i_2, ..., i_o\} = f_1(c_i) \cup f_2(c_i) \cup ... \cup f_n(c_i) \\ o = \sum_{x=0}^{n} |f_x(c_i)| \\ \Rightarrow \forall i_j \in I_i \exists f_k(c_i) \rightarrow i_j \in f_k(c_i), j \leq o \end{aligned}$$

Edges e_z are directed from the source node identified by u_a to a target node identified by u_b. They exist between any two nodes $v_a, v_b \in V$ that have a common identifier subset Λ_z so that $\Lambda_z \in I_a$ and $\Lambda_z \in I_b$ and furthermore no identifier has been matched in any other resource with an ordering index between t_a and t_b.

$$E = \{e_1, e_2, ..., e_q\}, e_z = (u_a, u_b, \Lambda_z),\ q, z \in N, z \leq q$$

$$\begin{gathered} \Lambda_z = \{i_r| \\ i_r \in I_a \wedge i_r \in I_b, \forall i_r \rightarrow V' = \{v_c|v_c = (u_c,t_c, I_c), i_r \in I_c \wedge t_a < t_c < t_b\} = \emptyset, \\ v_c \in V, r \in N, r \leq |I_b|\} \end{gathered}$$

The constructed networks of information co-occurrence are context-free in the sense that no global feature set or pre-existing structure is exploited for their generation, and no assumptions are made about the social networking architecture used. Edges result only from the comparison of pairs of resources. That does not mean (a) that no context exists, (b) that it is unimportant, or (c) that it should not be taken into account in the investigation of the cascade, only that we need to construct the cascade as an antecedent step, rather than a resulting one, because the structures we investigate will be biased if we weave assumptions about their context into their construction. By way of analogy, we might say: a forensic scientist will gather *all* evidence from a crime scene before formulating hypotheses about which evidence is relevant to the crime, as she should not begin with the assumption that Miss Scarlett did it with the lead piping, and then only collect the evidence that has a bearing on that narrower question. The wider set of evidence, lacking context, may at first appear incoherent, contradictory and coincidental, but once it is constructed it can be honed down with the reintroduction of context (e.g., evidence about someone with a watertight alibi can be discounted) to produce the specific narrative of the relevant explanans. For Transcendental Information Cascades, rich context can be added after the cascade construction, for example by weighting edges or nodes and modifying node labels.

The process described here yields different structures depending on both the data at hand and the information extraction algorithms applied, which serve to generate the particular cascade identifiers (i.e., identify information in the data set). Algorithms can be selected opportunistically, depending on (1) what is possibly significant, and (2) what structures are unlikely to be uncovered by more conventional methods; one could imagine searching for cascade identifiers within or among hashtags, URIs, quotes, topics, keywords, images, or even semantics and sentiments. Where appearances of information seem to co-occur serendipitously, we can focus our further investigation. Even though edges are only created between directly consecutive content elements that share an identifying pattern, implicitly any resource is in interaction – or entangled – with any other resource in the connected cascade it belongs to.

4 Studying socio-technical systems using Transcendental Information Cascades

We anticipate some general scenarios in studying socio-technical systems that would benefit from the application of Transcendental Information Cascades: the use of a single cascade to study a process; the use of multiple cascades to understand the significance of different types of information; and a framework to tie multiple cascades together into a coherent overall picture. We discuss these in turn.

4.1 Single cascades as a process: assessing intra-cascade properties

The nature of Transcendental Information Cascades – that they are directed networks preserving a concise set of informational patterns for each node – allows well-established quantitative methods to be used to capture structural as well as informational properties of socially constructed information traces. The benefit of this, even with only a single cascade, is that fundamental low-level analytical methods can be used, so that the system-specific context inherent to the analysis (e.g., case-specific feature sets) is reduced, allowing for unbiased discovery of significant patterns across (a cross-section of) socio-technical systems, rather than within (and therefore illicitly assuming the centrality of) particular restricted, well-behaved, and well-understood milieu. This allows determining where significant bursts of structural and/or informational patterns first appear or fade away, indicating the emergence or shift of an underlying exogenous phenomenon and providing a trigger for sampling a particular subset from the overall content (i.e., element sequence in a particular period of time that features a burst) for a detailed and further context-driven inspection.

4.2 Multiple cascades as a signpost of diversity: assessing inter-cascade properties

Given the intra-cascade properties identified above, can we expand an analysis to inter-cascade properties? In particular, given that different configurations of information extraction algorithms applied to the same sequence of content elements will result in differing cascade networks, one might reasonably ask: which of the resulting networks is the most *appropriate* representation of some underlying exogenous event or activity? Or phrased differently, one might ask: what is the appropriate information extraction algorithm configuration and process to best represent the underlying accumulated activities of human participants of the system?

An answer to this question would provide the basis for devising an adaptive approach to cascade construction. To start, entropy measures reflecting the distribution of cascade identifiers could be used to determine which matched informational patterns are associated with a certain degree of randomness. This then allows refinement of the information extraction algorithms by excluding certain patterns from consideration (e.g., specific words or hashtags used for spam on social media which tend to tie together information randomly, and thus spuriously, rather than reflecting populations priorities). Furthermore, bursty periods of different cascade networks can overlap (or be completely disjoint), indicating a relationship between (independence of) the extracted cascade identifiers. Where there is a relationship, we can then concatenate selected information extraction algorithms to construct another alternative cascade network.

4.3 Cohering multiple cascades: information in a multi-dimensional space

Detecting bursts of activity is a suitable means to infer exogenous events underlying socio-technical systems but it is typically focused on individual information streams (e.g., the recurrence of individual words over time without regarding their co-occurrence with other words) (Kleinberg, 2003; Barabasi, 2005). If we model cascades of information co-occurrence to describe the global interconnected informational state in a socio-technical system, we can represent information in a multi-dimensional space as shown in Figure **??** so that we can see bursts occurring along different axes.

Preserving the context-free nature of the approach, three dimensions may be the natural base for this representation: (1) time (or more generally, order in the set); (2) an index of all unique cascade identifier sets extracted from data (reflecting the chronological order in which identifier sets are found); (3) an index for each unique identifier set which is incremented with each occurrence of the respective set over time. Adding context allows us to scale the number of dimensions variably (e.g., adding further dimensions for the system in which particular information occurred or the human individual who shared it). It may be, for example, that we would want to include a geographical dimension, because we are interested in the specific viewpoints of the heterogeneous set of actors able to influence a situation (Cebrian et al., 2016); recall the events of the so-called ‘Twitter revolution’ in Iran in 2009, when thanks to over-reliance on the use of data from a single channel, Twitter the prospects of the revolutions success were dramatically over-estimated as most relevant Twitter traffic turned out to be supportive tweets from the US and the UK, and, as

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2789v2 | CC BY 4.0 Open Access | rec: 15 Apr 2018, publ: 15 Apr 2018

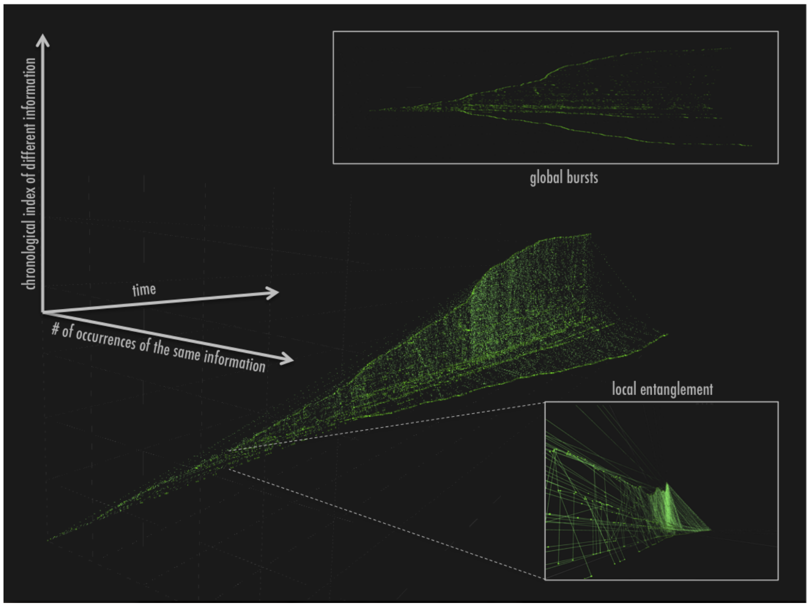

Fig. 3 Transcendental Information Cascades represented in a three dimensional space; (x) time; (y) information diversity as the chronological order in which unique identifier sets are found; (z) information specificity as the index for each unique identifier set.

(Honari, 2015) put it, ‘various areas of interest to Iranian users have been neglected or ignored’ in the literature.

This projection into a three-dimensional space allows the identification of (a) periods when new unique identifier sets are created at high frequency and (b) periods when particular identifier sets burst. While this space seems to naturally diverge over time from a macroscopic viewpoint, adding the cascade links to the visualization reveals ties across individual information streams allowing the tracing of time-persistent dependencies that would be otherwise hidden.

5 Applications of Transcendental Information Cascades

The motivating use case described earlier was the study of digital disaster response as a collective, opportunistic and improvisatory phenomenon, but there are many cases where a macroscopic view of the accumulated information sharing has more value than the architecturally amplified activities of individuals in a specific socio-technical system. Following guidelines for case-based research (Benbasat et al., 1987; Eisenhardt, 1989), we investigated the application of Transcendental Information Cascades to understanding real-world cases dependent on cross-channel communication: online citizen science and editing activities in Wikipedia.

5.1 Citizen science: coordination by content

Online citizen science is a blueprint of the trending hybrid computing approach, coupling state-of-the-art artificial intelligence with human computation, to enable interested people to tackle problems in scientific research that are impossible to solve in a purely computational fashion (Raddick et al., 2009). The Zooniverse, for example, is the world's largest multi-project citizen science platform, with over 1.3 million volunteers contributing to projects from various domains such as astrophysics, biology or digital humanities among others. The platform gained popularity as the source of numerous citizen-led discoveries made after participants had branched out beyond the immediate system-generated constraints, discussing outliers and making other remarkable serendipitous observations while performing the crowdsourcing task (Tinati et al., 2015b). Information sharing on those platforms often evolves to become domain-specific and goal-oriented. Hence, supporting this domain-specific information sharing around the objects examined as part of the crowdsourcing task has become part of the core of many citizen science systems. However, from the point of view of research methods in information dynamics, these systems are very peculiar with respect to the online communities they form. They typically do not feature explicit social networks and the community structures that emerge implicitly are highly fluid and dependent on many aspects of context (Luczak-Roesch et al., 2014).

Transcendental Information Cascades were applied to a dataset representing content sharing on the Planet Hunters project hosted on the Zooniverse (Luczak-Roesch et al., 2015b,c). Four different information extraction algorithms based on string matching using regular expressions were tested on this dataset in order to construct alternative cascade networks: (1) hashtags; (2) matching of content that refers to specific object identifiers related to the images shown in Planet Hunters; (3) matching of identifiers used by the Planet Hunters community to refer to objects in external astrophysics databases; (4) URIs. The studies of the resulting cascade networks as well as some of their network and information theoretic properties revealed that only the information extraction algorithms 2 and 3 were suited to be combined without further adaptation (see Figure 4). The cascades constructed by applying these methods naturally showed patterns of disjoint local phenomena, which were correlated in time. Meanwhile, cascades based on hashtags tended to be either single identifier cascades or consist of multiple roots that merged and diverged to form a single massive connected component from which little useful information could be extracted. URI-based cascade networks tended to feature a significant fraction of independent cascades in which one particular identifier set recurred repeatedly. Hence hashtag and URI cascades would need to be refined first, until the intra-cascade properties indicated the same distinctiveness as the other two approaches. Note that, in this case at least, the identifiers that were already built into the system were less insightful compared to expressions that evolved within the community (e.g. KID identifiers) and consistent with our argument to move beyond system-specific features to uncover interesting relationships.

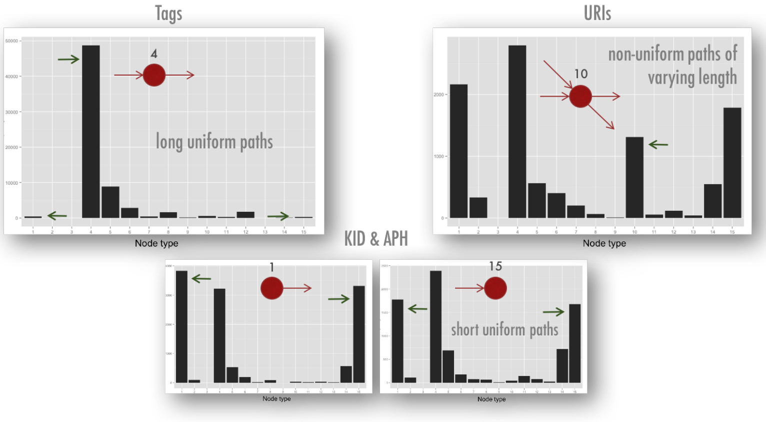

Fig. 4 The analysis of single node motifs shows the specific characteristic of the hashtag cascades (Tags). These feature a proportionally low number of network roots and stubs, which relates to many long uniform paths of hashtag recurrence that are not connected to any other matched hashtags. It is also worth highlighting the specific role of node type 10 in the URI cascades; this node type indicates that a unique matched set of URIs recurred frequently together with intermittent periods where subsets of those matched URIs occur independently. This figure was adapted from (**?**).

5.2 Wikipedia edits: a source of temporal relationships

Wikipedia represents a network of human-curated, moderated, and maintained articles, which over time have come to comprise the largest encyclopedia in existence. The variety of social processes in Wikipedia from managing vandalism, to ensuring quality and consistency in the knowledge base, and even to detecting gender imbalances allows us to consider it as more than just a network of explicitly linked articles. Implicit structures emerge from coordinated or sometimes just accumulated activities of volunteers, but can become explicit if the community approves them as useful, as exemplified by WikiProjects (Tinati et al., 2015a; Tinati and Luczak-Roesch, 2017), an effort to form sub-communities in order to increase the quality of domain-specific article sets. For Wikipedia, a core challenge is to discover and in certain situations support such emergent phenomena effectively within the vast amount of user and machine-generated data. Every second, hundreds of articles are created or revised, edits are overwritten or reverted, abuses are reported, and discussions take place. This stream of activities represents the digital traces of collective human action, and to that end, studying these streams reveals temporal relationships between articles that would otherwise remain hidden and promises to provide insight into the underlying social activities of such a system from a novel angle.

As an example of the potential for progression from context-free cascade construction to context enrichment for interpretation and sense making, we applied a string matching function to the text associated with each Wikipedia revision entry. The matching function uses a regular expression to identify trigram noun phrases to match entities like ‘The White House’, ‘Barack Hussein Obama’ or ‘Empire State Building’ for example. In this situation Transcendental Information Cascades form a network of article edits, linked together by the shared trigrams found within the edit revision text. By enriching the article edits with contextual knowledge about article categories from DBpedia (http://dbpedia.org) it was possible to find that this cascade network represents meaningful article relationships not available within the explicit network of linked Wikipedia articles (Tinati et al., 2016).

Further analysis of the informational and structural properties as well as the general burstiness characteristics (see Figure 5 of the constructed cascades showed that they reflect both external events and phenomena inherent to the system. In particular, a burst of activity was observed featuring a series of edits made within a short duration of time beginning with identifiers found in edits on the article about Edward Snowden. The cascade then branched out to span across many other articles incorporating various identifiers related to Edward Snowden’s life. A detailed inspection of the time frame when the cascade emerged showed that it coincided with a presentation given by him at the SXSW conference. In other words, a relationship between an external phenomenon and a short, bursty cascade of edits within Wikipedia, which would not have been available to a more contextualized investigation, was uncovered using the method.

In similar vein, we were also able to observe more local phenomena, such as a pathway found around the identifier: ‘U.S. District Court’. This cascade extended over a longer period of time, linking articles and identifiers related to same-sex marriage in the United States, which led to an editing debate within the Wikipedia community around articles featuring lists of U.S. state laws on same-sex unions. Here, in contrast to the cascade from Snowden’s talk, we were able to observe the frequent recurrence of articles within a single pathway indicating back-and-forth editing activity potentially even an edit war between Wikipedia editors. Any method that focuses either on individual users’ contributions over time or selected articles’ dynamics would have missed these article relationships by shared content that were made visible only through the application of TICs.

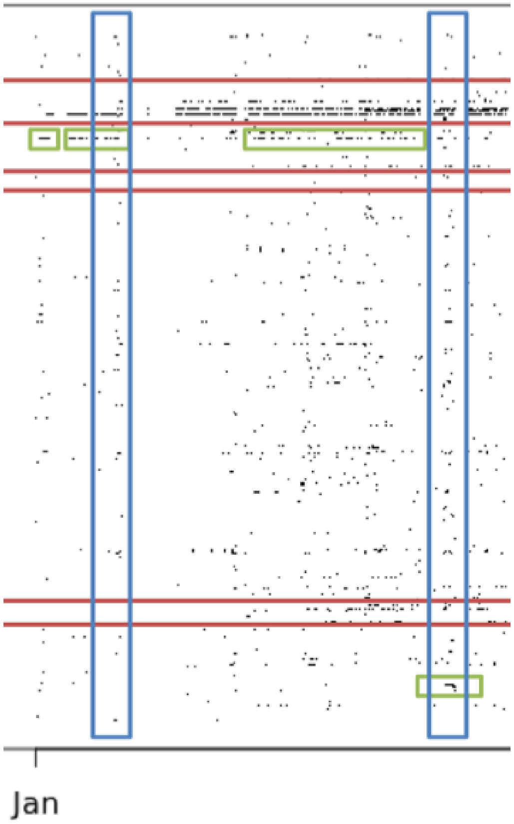

Fig. 5 Transcendental Information Cascades enabled a view that allows for differentiating between local bursts of individual trigrams (marked green), trigrams that feature continuously high editing activity as combined sets (marked red), and global bursts of activity that range across all matched trigrams (marked blue). This figure was adapted from (**?**).

6 Synthesizing the case of Transcendental Information Cascades

The applications of Transcendental Information Cascades described above show how their construction and analysis reveal hidden patterns of coordination within a stream of activity, focusing on information token co-occurrence independently from assumptions about causality, prior structure, and connectivity. This suggests that Transcendental Information Cascades may provide a unique way of underpinning the field of socio-technical systems with a distinct information-centric theory. Here we discuss a few attributes of TICs in favour of this suggestion.

6.1 Capturing objective knowledge about the macroscopic state of a socio-technical system

A common challenge in the representation and analysis of the state of socio-techical systems is the evaluation of the fitness of certain models that are the outcome of the cognitive process of a person performing the analysis. We suggest that this cognitive embedding situates those methods in 'world 2' according to Popper (Popper et al., 1972; Popper, 2013), who differentiated physical objects and events (world 1), mental objects and cognitive processes (world 2), and objective knowledge (world 3). Such models, situated in world 2, are not necessarily based on the same observational setup (e.g., the study of information diffusion based on modelling information cascades in a social network compared to the study of topic outbreaks modelled as activity bursts) and consequently are incommensurable views of the real state of a socio-technical system. Transcendental Information Cascades allow for an arbitrary amount of endogenous (e.g., structural properties of a particular TIC) or exogenous (e.g., features from the systems that emitted the data sequence that was transformed into that TIC) contextual features to be consulted for their analysis. However, time is the common dimension to all higher-dimensional views of different TICs of any given data sequence, and is always embedded into the network structure. Hence, we argue that time can be regarded as the only dimension independent of any context of analysis and, in consequence, independent of any cognitive model of reality imposed by the analyst. This situates Transcendental Information Cascades in world 3 according to Popper: objective knowledge about the state of a socio-technical system that can stand independently of any antecedent assumptions about what kind of network should be found (e.g., a social network is an imposed cognitive model of an online community's structure).

TICs may have application for the detection of influence of exogenous phenomena as well as temporal contagion within socio-technical systems, underlining that these contain social groupings, susceptible to influence from the full range of social contexts and social networks in which individuals take part, not simply the specific medium, platform, or architecture from which data can be harvested. In the case of Wikipedia in particular we also see great potential to uncover the injection of biases by focusing on information tokens rather than social features. We can expect people who perform such malicious editing to try to mask their identities and not to leave a digital trace allowing them to be detected and blocked. What they cannot mask is the information they inject and the TIC method allows this information to be traced over time.

To reiterate, none of this is meant to imply that data about, or gathered from, social networks is unimportant - far from it. But some extra input is required to understand what sort of intelligence is detectable within a socio-technical system as a whole, independently of assumptions about the mechanisms for its acquisition and delivery (and of course this independence is earned at the cost of restricting our use of assumptions about mechanism, at least as we construct the global information space).

6.2 Transcending the four views of information

A Transcendental Information Cascade can be regarded as a model that spans two of the four views of information presented by (McKinney Jr and Yoos, 2010). The low-level matching of patterns for cascade construction basically means looking for small but meaningful units in data sequences and reflects the ‘token view’ (McKinney Jr and Yoos, 2010; Lee, 2010). The mechanism to add relationships between ‘temporally coincident’ (Jung, 2010) occurrences of those tokens lets the model transcend to the higher-order ‘information in the syntax view’ (McKinney Jr and Yoos, 2010; Lee, 2010).

This step towards the syntax view suggests that a Transcendental Information Cascade channels and preserves information across time; specifically, it allows one to trace any captured information token back to the point in time when it was introduced into a socio-technical system and follow its path of co-occurrences with other tokens to the point when it eventually gets removed or replaced by another token. This is a unique feature of TICs, and implies that they have capacity to store and transfer information. Thus, TICs may be important for distributed communities, which may have few communally created information storage facilities capable of allowing access to information in a timely manner (e.g., at the point at which it is needed). Some, but not all, information that can be found right from the beginning when a socio-technical system is established remain present in the most recent state of that system, 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.

7 Conclusion

The aim of this work was to provide insight into a number of factors. First was the way in which the information flow is facilitated in socio-technical systems like the Web, abstracted away from the federation of co-created systems (and walled gardens). We argue that it is important to minimize the number of assumptions we make about the social context of information evolution not because we do not believe that social context plays a highly significant role, but rather to derive important social relationships from the information evolution, without reproducing the assumption that existing data from networking and sharing sites exhausts the relevant context (Facebook gives you the complete picture). This view provides an alternative, less powerful, less context-dependent, and potentially less deluded perspective on socio-technical systems in general and may be up for debate in the field as a kind of ‘box-breaking research’(Alvesson and Sandberg, 2014), raising a whole set of new questions about how we model and study emergent socio-technical systems.

Transcendental Information Cascades may be a complement to analyses that exploit rich contextual features as well as more complex *a priori* modelling or clustering of information (Shahaf et al., 2013). As various examples will show, sometimes the necessary contextual data is not available (or, under a different privacy regime, may not be accessible), in which case alternative techniques such as the ones proposed here would be required and welcome. The only general relationship presupposed by TICs is temporal precedence, and the key subjects of interest are *recurrence* (Eckmann et al., 1987; Webber Jr and Zbilut, 2005; Donner et al., 2011), or the return occurrence of information, and *bursts* (Kleinberg, 2003; Barabasi, 2005), or the multiple recurrences of some information at high frequency.

The transcendental understanding of cascades, following our Kantian theme (Kant, 1934), is skeptical about apparent causes and seeks their necessary conditions, rejecting the ready-made etiology contained in social network data and focusing instead on the narrower supporting base of whatever is detectable from time order and syntactic/semantic coincidence. Our attempt to devise an information-centric theory for socio-technical systems enables a macroscopic view of the emergent output of complex social action by studying the change of network and information theoretic properties, which suggests a link to social entropy theory (Bailey, 1990, 2006). This differentiates Transcendental Information Cascades from the system-centric perspectives commonly referred to in Social Computing and Computer-supported Cooperative Work (Grudin, 1994; Parameswaran and Whinston, 2007).

We hope to be able to understand any kind of socio-technical systems as a wider phenomenon than a siloed representation in a specific network might imply, a coherent phenomenon, a chord rather than its arpeggiated components, expressing its state at a time by quantifying the information represented (and its dynamics). This is valuable for research on Social Machines (Berners-Lee et al., 2000; Hendler and Berners-Lee, 2010), as characterized in (Smart et al., 2014) as 'Web-based socio-technical systems in which the human and technological elements play the role of participant machinery with respect to the mechanistic realization of system-level processes'. Our work contributes insight into the organic 'system-level processes', so that the computation of Social Machines becomes the output of this kind of analysis (Luczak-Roesch et al., 2015a), rather than one of the inputs, and no assumptions are made about Social Machines as marooning themselves on particular channels (on which we happen to have the data). This suggests that there exists an interesting new generic phenomenon in socio-technical systems that we call *not necessarily coordinated collectives*. **A not necessarily coordinated collective is a group of people treated as equal contributors to an accumulated activity stream, regardless of any pre-defined real or virtual relationships between those people or their contributed content.**

The models and experiments we have discussed here are of course very small steps on what will necessarily be a long journey of research, experimentation, and much more complex macroscopic and microscopic investigation. For instance, how do we determine the relationship between all possible Transcendental Information Cascades of a given information sequence; find the best partitioning of a cascade network into the minimum number of non-nested substructures to derive an aggregated state machine representation; or mine the collective intent of the people involved in the contents of particular Transcendental Information Cascades? It needs to be investigated whether this novel view allows for predictions of aspects of the underlying system that are at par or better than those of highly contextualized methods. And we need the capacity to index and search for not only documents and data as for classical Web graphs (Broder et al., 2000) but also Transcendental Information Cascades themselves, enabling us to understand how information dynamics facilitate and are facilitated by procedural knowledge. In the end, such understanding will have engineering repercussions, as we seek to create the conditions for the effective creation of knowledge.

TICs require extensive measurement and understanding, but many of the tools are readily at hand. They are a transformation of any kind of source data into the same well-defined temporal network model, which adds mature methods from network analysis to the analytical toolbox (Holme and Saramäki, 2012; Williams and Musolesi, 2016). And when TIC paths collide at a particular point in time, the tools of information theory (Shannon, 1949; Kullback, 1997) can be used to understand the properties of the collision and the nature of the resulting entanglement as entropy will increase or decrease. Finally, motifs of the network structure or the entropy over time can be aggregated into states, which have their own analytical apparatus grounded in stochastic processes (Anick et al., 1982; Parisi and Sourlas, 1982; Rabiner, 1989; Ovchinnikov, 2016) but also in nonlinear dynamic systems theory (Strogatz, 2014).

Beyond the application of Transcendental Information Cascades to socio-technical systems we discussed here, TICs are a generic method that makes formerly hidden dimensions of any time-evolving system visible. We also explore the application of TICs to historic corpora of literature, EEG brain wave recordings, and gene sequences for example. The method always leads to the same kind of temporal network, independent from the type of input data and the analytical context, and time becomes the one dimension that is common to all possible views to the Transcendental Information Cascade in higher-dimensional space representations. This allows for seeing temporal dynamics that have not been accessible before in a unified way, and leads to novel questions about emergence, chaos, and randomness.

References

Adamic, L. A. and Huberman, B. A. (2000). Power-law distribution of the world wide web. *science*, 287(5461):2115–2115.

Alvesson, M. and Sandberg, J. (2014). Habitat and habitus: Boxed-in versus box-breaking research. *Organization Studies*, 35(7):967–987.

Anick, D., Mitra, D., and Sondhi, M. M. (1982). Stochastic theory of a data-handling system with multiple sources. *Bell Labs Technical Journal*, 61(8):1871–1894.

Bailey, K. D. (1990). *Social entropy theory*. SUNY Press.

Bailey, K. D. (2006). Living systems theory and social entropy theory. *Systems Research and Behavioral Science*, 23(3):291–300.

Barabasi, A.-L. (2005). The origin of bursts and heavy tails in human dynamics. *arXiv preprint cond-mat/0505371*.

Barabási, A.-L., Albert, R., and Jeong, H. (2000). Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: statistical mechanics and its applications*, 281(1-4):69–77.

Benbasat, I., Goldstein, D. K., and Mead, M. (1987). The case research strategy in studies of information systems. *MIS quarterly*, pages 369–386.

Berners-Lee, T., Fischetti, M., and Foreword By-Dertouzos, M. L. (2000). *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. HarperInformation.

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., and Wiener, J. (2000). Graph structure in the web. *Computer networks*, 33(1):309–320.

Cebrian, M., Rahwan, I., and Pentland, A. S. (2016). Beyond viral. *Communications of the ACM*, 59(4):36–39.

Donner, R. V., Small, M., Donges, J. F., Marwan, N., Zou, Y., Xiang, R., and Kurths, J. (2011). Recurrence-based time series analysis by means of complex network methods. *International Journal of Bifurcation and Chaos*, 21(04):1019–1046.

Eckmann, J.-P., Kamphorst, S. O., and Ruelle, D. (1987). Recurrence plots of dynamical systems. *EPL (Europhysics Letters)*, 4(9):973.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of management review*, 14(4):532–550.

Grudin, J. (1994). Computer-supported cooperative work: History and focus. *Computer*, 27(5):19–26.

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

Holme, P. and Saramäki, J. (2012). Temporal networks. *Physics reports*, 519(3):97–125.

Honari, A. (2015). Online social research in iran: A need to offer a bigger picture. *CyberOrient: The Online Journal of Virtual Middle East*, 9(2).

Jung, C. G. (2010). *Synchronicity: An Acausal Connecting Principle.(From Vol. 8. of the Collected Works of CG Jung)(New in Paper)*. Princeton University Press.

Kant, I. (1934). *Critique of Pure Reason. Translated by Norman Kemp Smith.* London Macmillan 1934.

Keegan, B. C., Lev, S., and Arazy, O. (2016). Analyzing organizational routines in online knowledge collaborations: A case for sequence analysis in cscw. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pages 1065–1079. ACM.

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

Kraut, R. E., Resnick, P., Kiesler, S., Burke, M., Chen, Y., Kittur, N., Konstan, J., Ren, Y., and Riedl, J. (2012). *Building successful online communities: Evidence-based social design*. Mit Press.

Kullback, S. (1997). *Information theory and statistics*. Courier Corporation.

Lee, A. S. (2010). Retrospect and prospect: information systems research in the last and next 25 years. *Journal of Information Technology*, 25(4):336–348.

Lee, C. P. and Paine, D. (2015). From the matrix to a model of coordinated action (moca): A conceptual framework of and for cscw. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 179–194. ACM.

Lerman, K. (2016). Information is not a virus, and other consequences of human cognitive limits. *Future Internet*, 8(2):21.

Lerman, K., Yan, X., and Wu, X.-Z. (2016). The" majority illusion" in social networks. *PloS one*, 11(2):e0147617.

Luczak-Roesch, M., Tinati, R., O'Hara, K., and Shadbolt, N. (2015a). Socio-technical computation. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*, pages 139–142. ACM.

Luczak-Roesch, M., Tinati, R., and Shadbolt, N. (2015b). When resources collide: Towards a theory of coincidence in information spaces. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1137–1142. ACM.

Luczak-Roesch, M., Tinati, R., Simperl, E., Van Kleek, M., Shadbolt, N., and Simpson, R. J. (2014). Why won't aliens talk to us? content and community dynamics in online citizen science. In *ICWSM*.

Luczak-Roesch, M., Tinati, R., Van Kleek, M., and Shadbolt, N. (2015c). From coincidence to purposeful flow? properties of transcendental information cascades. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, pages 633–638. ACM.

Malone, T. W., Laubacher, R., and Dellarocas, C. (2009). Harnessing crowds: Mapping the genome of collective intelligence.

McKinney Jr, E. H. and Yoos, C. J. (2010). Information about information: A taxonomy of views. *MIS quarterly*, pages 329–344.

McLuhan, M. and Fiore, Q. (1967). The medium is the message. *New York*, 123:126–128.

Ovchinnikov, I. V. (2016). Introduction to supersymmetric theory of stochastics. *Entropy*, 18(4):108.

Parameswaran, M. and Whinston, A. B. (2007). Research issues in social computing. *Journal of the Association for Information Systems*, 8(6):336.

Parisi, G. and Sourlas, N. (1982). Supersymmetric field theories and stochastic differential equations. *Nuclear Physics B*, 206(2):321–332.

Popper, K. (2013). *Knowledge and the Body-Mind Problem: In defence of interaction*. Routledge.

Popper, K. R. et al. (1972). Objective knowledge: An evolutionary approach.

Quinn, A. J. and Bederson, B. B. (2011). Human computation: a survey and taxonomy of a growing field. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1403–1412. ACM.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

Raddick, M. J., Bracey, G., Carney, K., Gyuk, G., Borne, K., Wallin, J., Jacoby, S., and Planetarium, A. (2009). Citizen science: status and research directions for the coming decade. *AGB Stars and Related Phenomenastro 2010: The Astronomy and Astrophysics Decadal Survey*, 2010:46P.

Shahaf, D., Yang, J., Suen, C., Jacobs, J., Wang, H., and Leskovec, J. (2013). Information cartography: creating zoomable, large-scale maps of information. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1097–1105. ACM.

Shannon, C. E. (1949). Communication theory of secrecy systems. *Bell Labs Technical Journal*, 28(4):656–715.

Smart, P., Simperl, E., and Shadbolt, N. (2014). A taxonomic framework for social machines. In *Social Collective Intelligence*, pages 51–85. Springer.

Strogatz, S. H. (2014). *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Hachette UK.

Tarski, A. (1944). The semantic conception of truth: and the foundations of semantics. *Philosophy and phenomenological research*, 4(3):341–376.

Tinati, R. and Luczak-Roesch, M. (2017). Wikipedia: a complex social machine. *SIGWEB Newsletter*, pages 1–10.

Tinati, R., Luczak-Roesch, M., and Hall, W. (2016). Finding structure in wikipedia edit activity: An information cascade approach. In *Proceedings of the 25th International Conference Companion on World Wide Web*, pages 1007–1012. International World Wide Web Conferences Steering Committee.

Tinati, R., Luczak-Roesch, M., Shadbolt, N., and Hall, W. (2015a). Using wikiprojects to measure the health of wikipedia. In *Proceedings of the 24th International Conference on World Wide Web*, pages 369–370. ACM.

Tinati, R., Van Kleek, M., Simperl, E., Luczak-Rösch, M., Simpson, R., and Shadbolt, N. (2015b). Designing for citizen data analysis: A cross-sectional case study of a multi-domain citizen science platform. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 4069–4078. ACM.

Webber Jr, C. L. and Zbilut, J. P. (2005). Recurrence quantification analysis of nonlinear dynamical systems. *Tutorials in contemporary nonlinear methods for the behavioral sciences*, pages 26–94.

Whitehead, A. N. and Russell, B. (1912). *Principia mathematica*, volume 2. University Press.

Williams, M. J. and Musolesi, M. (2016). Spatio-temporal networks: reachability, centrality and robustness. *Royal Society open science*, 3(6):160196.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., and Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *science*, 330(6004):686–688.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2789v2 | CC BY 4.0 Open Access | rec: 15 Apr 2018, publ: 15 Apr 2018