# Modeling Information Diffusion in Social Media as Provenance with W3C PROV

Io Taxidou[*]          Tom De Nies[‡]          Ruben Verborgh[‡]

Peter M. Fischer[*]          Erik Mannens[‡]          Rik Van de Walle[‡]

[*]{taxidou,peter.fischer}@informatik.uni-freiburg.de
University of Freiburg
Germany
[‡]{tom.denies,ruben.verborgh,erik.mannens,rik.vandewalle}@ugent.be
Ghent University - iMinds - Multimedia Lab
Belgium

## ABSTRACT

In recent years, research in *information diffusion* in social media has attracted a lot of attention, since the produced data is fast, massive and viral. Additionally, the *provenance* of such data is equally important because it helps to judge the relevance and trustworthiness of the information enclosed in the data. However, social media currently provide insufficient mechanisms for provenance, while models of information diffusion use their own concepts and notations, targeted to specific use cases. In this paper, we propose a model for information diffusion and provenance, based on the *W3C PROV Data Model*. The advantage is that PROV is a Web-native and interoperable format that allows easy publication of provenance data, and minimizes the integration effort among different systems making use of PROV.

## 1. INTRODUCTION

Social media such as online social networks (e.g. Facebook), micro-messaging services (e.g.Twitter) or sharing sites (e.g. Instagram) provide the virtual space in which a significant part of social interactions takes place. Many real-life situations, such as elections, are reflected by social media. In turn, social media shape these situations by forming opinions or strengthening trends, or by spreading reports on emerging situations faster than conventional media. Furthermore, word of mouth plays an important role in shaping user's attitudes and behavior. Most importantly, social media provide a huge audience (some users maintain millions of connections) where information can be easily spread and consumed by others. This phenomenon is referred as *information diffusion*.

Because there exists a plurality of opinions and multiple sources of information in social media, the need for judging the relevance and trustworthiness of such information is becoming urgent. The understanding of how a piece of information propagated in social media provides additional context, including the source and its properties, the intermediate forwarders and the modifications that
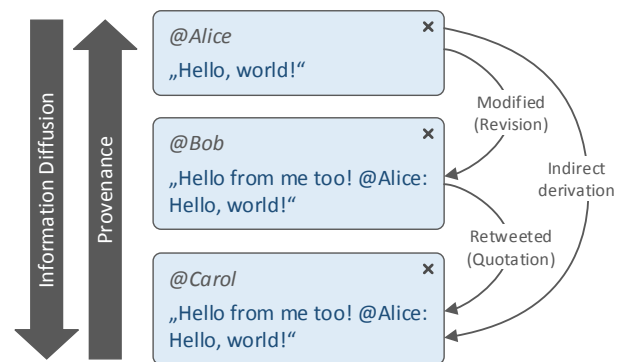
Figure 1: Information Diffusion and Provenance

this piece of information has undergone. A social media user can take advantage of this context to assess how much value, trust, and validity such information carries.

For example, online journalists need to understand the cycle of information diffusion in a timely manner, by assessing the source and intermediate forwarders, predicting information *virality* as well as determining the impact of their own publications. Additionally, the detection of rumors is feasible not only by discovering the sources but also by analyzing the properties of the diffusion process [13] and the intermediate steps. When it comes to massive amounts of negative opinions expressed in social media, companies, politicians and celebrities need to react promptly by understanding who is propagating certain information and who is influencing others.

Such kind of analysis refers to the reverse process of information diffusion, *information provenance*, that seeks the paths back to the sources. While provenance is a well researched topic in domains like workflows [9] or databases [5], it has received limited attention in the context of social media, compared to classical information diffusion. Likewise, existing models of information diffusion are insufficient to model provenance, while the current structure of social media provides limited or no mechanism to its users to judge received information [4]. For example, in cases of retweets on Twitter, only the source of information is provided but not the intermediate steps (forwarders). However, it has been shown that forwarders play an equally important role in the outcome of information diffusion [3].

To further clarify the relation between information diffusion and provenance, we provide an example in Figure 1. Three Twitter users are emitting a similar message: Alice is the source of information diffusion, as she emits an original message. At a later point in time, user Bob modifies the original message and then user Carol copies and forwards (retweets) the message of Bob. In this process, it is important to understand how the message was modified and forwarded. User Carol was indirectly influenced by user Alice, since her message was *indirectly derived* from the source (two-step procedure). This means that the trustworthiness of all three users involved should be judged, since they participate in the diffusion and modification of this message.

Despite the variety of models of information diffusion, there is currently no unified, conceptual model for information diffusion and provenance that can be applied to different datasets and set-ups, while remaining both expressive and generic enough to cover many use cases. In this paper we provide such a conceptual model. More specifically, we provide **PROV-SAID**, a model to assert the **Prov**enance of **S**ocial medi**A** **I**nformation **D**iffusion based on PROV-DM [17]. PROV-DM is the main component of the wider family of W3C PROV documents and defines a model for provenance. The PROV specification provides the concepts and supporting definitions to enable the inter-operable interchange of provenance information in heterogeneous environments such as the Web.

PROV-DM has many benefits; concepts of information diffusion and provenance are modeled in a W3C standard that allows subsequent integration and interaction with other tools that make use of PROV. As a result, the cost of integration is lowered since data derived from this modeling are exposed in a Web-native and interoperable format. This is very useful in cases where data needs to be combined from different (social media) sources that do not share the same concepts and notations. Additionally, PROV-DM is domain-agnostic, but it has the benefit of extensibility, allowing domain-specific information to be included.

As our main contribution in this paper, we introduce a number of new attribute values to extend PROV-DM [17], and relevant extensions to PROV-Constraints [7] to govern the use of these attributes values. In more detail, we provide: (1) a structured ontology for information diffusion and provenance on social media; (2) extensions of entities and activities relevant for information diffusion and provenance; (3) introduction of the use of these new concepts as attributes and roles in PROV assertions; (4) extensions for the vaguely defined concept of *Influence* in PROV-DM. Note that on the one hand, our model allows the representation of social connections among users, since information flows through them in the majority of cases. On the other hand, the model is generic enough to assert the provenance of information diffusion even without the presence of social connections.

The rest of the paper is structured as follows: in Section 2 we describe related work, in Section 3 our model, extensions and relevant constraints and in Section 5 we elaborate into the concept of Influence. Lastly, Section 6 concludes our paper.

## 2. RELATED WORK

In this interdisciplinary work, we can discern three categories of related work: (i) models of information diffusion and provenance in social media, (ii) provenance use cases for the PROV-DM model, and (iii) extensions to improve the expressiveness of the model itself or to cater for specific use cases.

Information Diffusion in social media and networks has been a well researched topic. A review of relevant models can be found in [10]. Until now, the focus has been on the design of

models with specific goals [4], e.g. assessing the probability that certain users are being reached). Such research is mostly driven by data mining techniques, (algorithms, frameworks and systems) to analyze specific datasets. This sort of analysis is useful in use cases such as marketing, solving the problem of maximizing the spread of information by targeting a specific users (i.e., the *influence maximization* problem [12]).

While provenance is a thoroughly investigated topic in other domains [5, 9], existing models of information diffusion do not provide the means to express it. A review of challenges and methods for provenance on social media can be found in [4]. Authors propose their own method inducing both user attributes and network structure. Our work is complementary, since we provide a general model for provenance to be used in different use cases and they propose certain metrics and algorithms to assess provenance. An example of using simple user attributes (e.g. authority score) to express provenance is the Twitcident system [1], that traces emergent events in Twitter. However, information concerning modifications that tweets undergo and intermediate steps, is not being exposed. The work of [19] and [20] presents a system and visualizations for reconstructing diffusion paths in real-time on social media. The proposed algorithm searches for all possible diffusion paths back to the sources and offers the possibility of different influence models in case it is not clear which paths the information took.

PROV-DM has been applied to various use cases. For example, authors in [16] used this model to express history of revisions in Wikipedia. In a different domain, authors model the history of clinical guidelines with PROV [14]. Such work facilitates the understanding of provided recommendations by practitioners. More use cases are listed in the PROV implementation report [11], and the list continues to grow.

PROV was deliberately kept as generic and extensible as possible, to allow for all possible use cases. For example, in the context of neuroimaging, the PROV-DM was extended in order to capture provenance between the stages that neuroimaging data undergoes: data acquisition, pre-processing and statistical analysis [15]. This way, relevant parameters used in each stage can be traced back which facilitates reproducibility and metadata analysis. A general extension to PROV-DM was proposed by [8] in order to capture the concept of uncertainty in two ways: uncertainty in provenance statements and uncertainty about the content of an entity whose provenance is assessed. This last extension is very useful when algorithms with a certain degree of uncertainty are used to assert the provenance. For example, if an information diffusion detector were used to assert the PROV graphs mentioned throughout this paper, UP [8] would be a way to annotate the provenance with the detector's confidence.

## 3. MODEL OVERVIEW

In Sections 3, 4 and 5, we describe our model with its relevant extensions and constraints[1]. PROV has formal semantics [6], which cover our model as well, since our extensions and constraints are fully compliant with PROV.

Throughout the text, we provide a full example that covers all aspects of the model. To improve clarity, this example is unfolded incrementally and the reader should take into account information in previous examples.

## 3.1 Overview

---

[1]For detailed specification and formal constraints, see `http://semweb.mmlab.be/ns/prov-said/PROV-SAID.html`
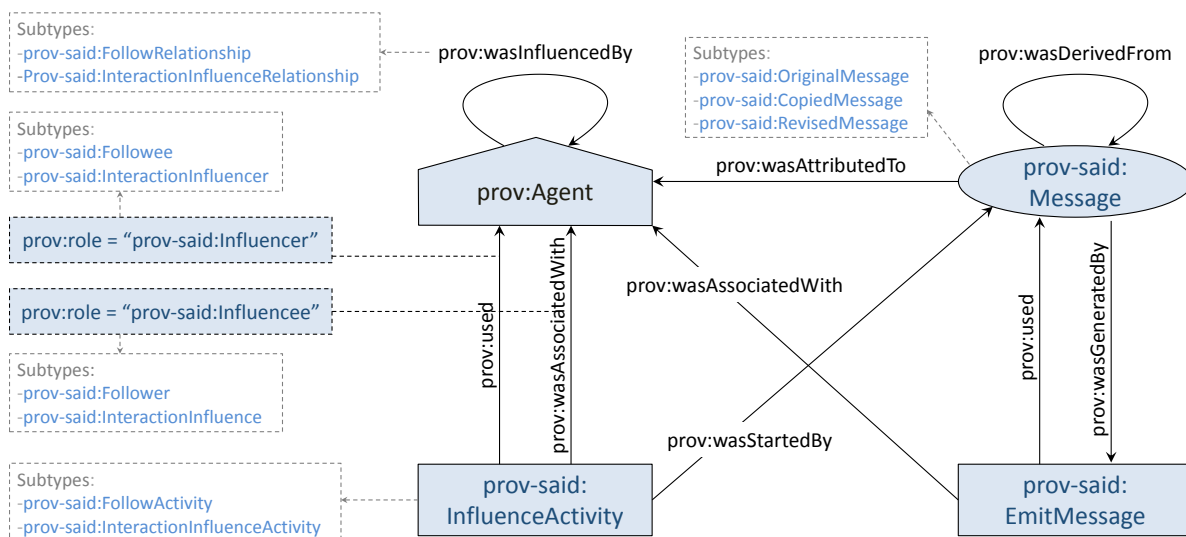
Figure 2: Prov-SAID Model

The PROV-SAID model can be applied to any social network where information propagates from user to user in the form of messages. Messages can be transmitted though *social connections*, but the model is general enough to capture *external influence* as well, as often happens in social media [18]. For example, Twitter users might publish information that has been seen on the public timeline without any direct social connection. Furthermore, our experience with provenance in Twitter shows that information does not flow only from social connections, but there is approximately 20% of external influence [19]. The last observation derives from experiments with reweets where diffusion is explicit, while this percentage is much higher for non explicit diffusion (propagation of Twitter hashtags).

Our model includes *activities* and *relationships* connected with information diffusion, such as exchanging *messages*, finding the *source* of diffusion, and expressing which *changes* the message has undergone through this procedure. User *influence* plays a key role in information diffusion since it drives information flow. The concept of influence in PROV-DM is vaguely defined and it is recommended to use more specific terms when possible. This is sensible since influence can take many forms in different use cases. However, for our use case the influence relationship has its own merit. Therefore, we define and extend the concept of influence, expressed through different activities, types and user roles.

Figure 2 shows a high-level overview of the PROV-SAID model. The proposed extensions to the standard are written in a blue font. In the next sections, we will describe each component in detail. Throughout this description, the prefix *prov:* refers to the PROV namespace[2] and the prefix *prov-said:* refers to the new PROV-SAID namespace[3]. Users who emit messages on social media are represented by the *prov:Agent* concept.

## 3.2 Design Decisions

The purpose of PROV-SAID is to offer an easily reusable model that covers and infers different aspects of information diffusion and provenance. Note that the goal is not minimizing the relationships in the model, but offering maximum expressiveness.

Also, we aim at borrowing the already defined concepts from PROV-DM wherever possible, and defining our own extensions for specific use cases. This way we improve clarity and we encourage reusability of the model. One example of extending the model, is the concept of *prov:Influence*: We differentiate the cases in the context of information diffusion and provenance and we give a more clear meaning to them.

The PROV-SAID model refers to information diffusion and provenance in the context of social media; since social connections are the main carriers of information [19] we need to specify whether a message was propagated through them or whether there was some external influence. For this purpose, we implicitly model social graphs connections (unidirectional relationships) produced through *follow* activities. We proceed by describing the components and relationships of our model step by step.

## 4. MODELING MESSAGES

In order to model messages that are emitted by users, we propose the following extensions that are subtypes of *prov:Entity*:

- *prov-said:Message*: denotes the general class of messages. Messages in social media might be original messages, copied messages or revised messages We define the following categories as subtypes of *prov-said:Message*:
- *prov-said:OriginalMessage* denotes an original message that is not derived from any other message and the user who emitted it is the initiator of information propagation for a specific message.
- *prov-said:CopiedMessage* denotes a message which is based on another message that has been published in the past and was forwarded as an exact copy. Users who emit copied messages comply fully with the content and opinions of the original message. For example, Twitter offers the *retweet* function were users can easily forward copies of messages emitted by others.
- *prov-said:RevisedMessage* denotes a message that is produced by modifying an existing message. This means that the user who emits such a message may or may not share the original opinion of the original message. It is possible that the information carried by the original message is altered.

With these three cases, we have covered the main cases of information diffusion through messages. This model reflects both the

---

[2]http://www.w3.org/ns/prov#
[3]http://semweb.mmlab.be/ns/prov-said/

reality of message-based information diffusion in social media and the conceptual foundations established in the research on database provenance as *provenance models* [5].

## 4.1 Message Attribution

A *prov-said:Message* is always attributed to a user *prov:Agent* using the relationship *prov:wasAttributedTo*. Example 1 illustrates the use of messages and attribution for the Twitter social network.

*Example 1: Message Creation and Attribution.*

```
prefix twitter: <http://twitter.com/>
prefix alice-status:
   <http://twitter.com/Alice/status/>
prefix bob-status:
   <http://twitter.com/Bob/status/>
prefix carol-status:
   <http://twitter.com/Carol/status/>

// User @Alice tweeted a message "Hello, world!"
prov:entity(alice-status:12345,
   [prov:type='prov-said:OriginalMessage',
   prov:label='Hello, world!'])

// User @Bob modified and re-emitted the
// "Hello, world!" message
prov:entity(bob-status:23456,
   [prov:type='prov-said:RevisedMessage',
   prov:label='Hello from me too!
   MT @Alice: Hello, world!'])

// User @Carol retweeted (copied)
// the revised message
prov:entity(carol-status:67891,
   [prov:type='prov-said:CopiedMessage',
   prov:label='Hello from me too!
   MT @Alice: Hello, world!'])

// alice-status:12345 was emitted
// by twitter:Alice
prov:wasAttributedTo(alice-status:12345,
   twitter:Alice)
```

## 4.2 Message Emission

Next we define the following activity that refers to message emission and is a subtype of *prov:Activity*

- *prov-said:EmitMessage* denotes a generic emission of a message. It must generate a *prov-said:Message*, and may use another *prov-said:Message*.

Note that the subtype of the generated *prov-said:Message* (original, copied or revised) can be inferred from the usage of another *prov-said:Message* by the *prov-said:EmitMessage*. If the content of the generated message is identical to that of the used one, it is a *prov-said:CopiedMessage*. If the content of the generated message was altered from that of the used one, it is a *prov-said:RevisedMessage*.

## 4.3 Message Derivation

Whereas an original message does not have dependencies on other messages, copied and revised messages can be traced back to their original sources through derivation. PROV-DM already provides most of the concepts needed to model this, in the form of *prov:Quotation*, *prov:Revision*, and *prov:PrimarySource*, as illustrated by Example 2.

*Example 2: Message Derivation.*

```
// bob-status:23456 was derived from
// alice-status:12345, which is also its
// primary source (in the context of Twitter)
prov:wasDerivedFrom(bob-status:23456,
   alice-status:12345, emit-23456, gen-23456,
   use-12345, [prov:type='prov:Revision',
   prov:type='prov:PrimarySource'])


// carol-status:67891 was quoted from
// bob-status:23456
// (which is not its primary source)
prov:wasDerivedFrom(carol-status:67891,
   bob-status:23456, emit-67891, gen-67891,
use-23456, [prov:type='prov:Quotation'])
```

We observe that *carol-status:67891* was indirectly derived from *alice-status:12345*. To model this dependency, we introduce the concept *prov-said:IndirectDerivation*. This way we can model multi-step provenance and trace how messages are being derived, without being restricted to the previous step only. We illustrate this in Example 3.

*Example 3: Indirect Derivation.*

```
// carol-status:67891 was indirectly
// derived from alice-status:12345
prov:wasDerivedFrom(carol-status:67891,
   alice-status:12345,
   [prov:type='prov-said:IndirectDerivation'])
```

At this point, we express the following constraints:

- An *prov-said:OriginalMessage* cannot be derived from a *prov-said:Message*.
- A copied or revised message should always be derived from another message. A *prov-said:EmitMessage* that generates a *prov-said:CopiedMessage* and uses a *prov-said:Message* implies that the first message was derived from the latter by means of *prov:Quotation*. Analogously, generation of a *prov-said:RevisedMessage* and usage of a *prov-said:Message* by a *prov-said:EmitMessage* implies that the first message was derived from the later by *prov:Revision*.

In the next Section we continue with modeling Influence with regard to information diffusion.

## 5. MODELING INFLUENCE

Influence plays a key role since it drives information flow in social media; however, it is hard to capture influence both from its semantics and sources. Users are often being influenced by external factors such as traditional media and react in social media [18]. We model influence in the closed world of social media. Influence has two ways of being expressed: through establishing social connections and through exchanging messages. Influence in PROV-DM is vaguely defined and we need a more expressive modeling to capture its different forms. We propose extensions for influence types, influence activities and influence roles.

## 5.1 Influence types

We define a relationship: *prov-said:InfluenceRelationship* to express general influence and two subtypes to specify the ways that such influence can be expressed.

- *prov-said:InfluenceRelationship* is a subtype of *prov:Influence*. It denotes an influence between agents in the context of social media. We define the following two subtypes:

- *prov-said:FollowRelationship* denotes that one agent was influenced by another agent, by establishing a unidirectional (follow) relationship. In the context of social media, that practically means being exposed to the messages emitted by the latter. For example, in Twitter this is the only way of connecting with users, while Facebook apart from the bidirectional *Friendship*, also gives the possibility of unidirectional connection by subscribing to the messages of users. Here, we assume that once an agent starts to follow another agent, he is exposed to his old messages and his future ones (if there are any). This is also the case in Twitter and Facebook. As a result, we do not model the influence that derives from exposure/subscription to messages explicitly, since it is implied by the *prov-said:FollowRelationship*. Furthermore, such a relationship entails a certain degree of uncertainty, since it can not be asserted whether an agent has seen the messages of another in reality.
- *prov-said:InteractionInfluenceRelationship* is a subtype of *prov-said:InfluenceRelationship*. It denotes that an agent was influenced by another agent, by having quoted or revised messages of the latter. As mentioned in Section 4.2, such a relationship can be discovered by investigating the similarity of messages of the former to the latter.

We illustrate the influence types in Example 4.

### Example 4: Influence Types.

```
// User @Alice followed user @Carol, so a
// prov-said:FollowRelationship existed between
// them
prov:wasInfluencedBy(twitter:Alice,
   twitter:Carol,
   [prov:type='prov-said:FollowRelationship'])

// User @Bob revised a message from
// @Alice, so a
// prov-said:InteractionInfluenceRelationship
// existed between them
prov:wasInfluencedBy(twitter:Bob,
   twitter:Alice,
   [prov:type='prov-said:
      InteractionInfluenceRelationship'])
```

By following these influence types, both the social graph and the interaction graph [21] can be reconstructed at a certain point in time by using provenance. The interaction graph aggregates interactions (e.g. emission of messages) among users as weighted edges.

## 5.2 Influence Activities

Additionally to the influence types expressed as relationships (subtypes of *prov:Influence*) among users provide modeling for influence, we explicitly model the corresponding activities. This design decision offers greater expressiveness by providing more information about their start and end time, what triggered them etc. For these purposes, we introduce three subtypes of *prov:Activity*:

- *prov-said:InfluenceActivity* is subtype of prov:Activity. It denotes the activity of one agent influencing another with the following two subtypes:
- *prov-said:FollowActivity* denotes the activity of one agent to establish a unidirectional connection with another. Once such an activity starts, the first agent is exposed to the (future and past) message emissions of the later. This activity has a start time that denotes the time of establishing the connection and an optional end time in case the agent removes the connection with regard to the other agent.

- *prov-said:InteractionInfluenceActivity* denotes the activity of one agent to influence another, so that the latter interacts by forwarding the messages of the first. Note here that the *prov-said:InteractionInfluenceActivity* is instantaneous, and thus has the same start and end time. In this way, we are able to model multiple interactions of agents by generating multiple instances of *prov-said:InteractionInfluenceActivity*. If we had considered the opposite case when *prov-said:InteractionInfluenceActivity* is asserted only once and has no end time, we would have come to contradiction with the principles of information diffusion, where the significance of past interactions fades quickly over time.

Example 5 illustrates the subtypes of *prov-said:InfluenceActivity*.

### Example 5: Influence Activities.

```
// A prov-said:FollowActivity was started at the
// moment user @Alice followed user @Carol.
// Since @Alice was still following @Carol
// at the time of assertion, there is no end time
// for the activity.
activity(alice-follows-carol,
   2015-01-09T13:00:00, - ,
   [ prov:type='prov-said:FollowActivity'])

// A prov-said:InteractionInfluenceActivity was
// started at the moment user @Bob
// modified and re-emitted the message of @Alice.
activity(bob-influencedby-alice,
   2015-01-09T13:05:00, 2015-01-09T13:05:00,
   [prov:type='prov-said:
      InteractionInfluenceActivity'])
wasStartedBy(bob-influencedby-alice,
   bob-status:23456, emit-23456,
   2015-01-09T13:05:00)
wasEndedBy(bob-influencedby-alice, bob-status:23456,
   emit-23456, 2015-01-09T13:05:00)
```

## 5.3 Influence Roles

Analyses of information diffusion and influence in social media make use of specific roles for their agents [2]. To model this, we need to specifically define values for the *prov:role* attribute in the context of *prov:Usage* and *prov:Association*. This way, we clarify the roles of agents involved in a *prov-said:InfluenceActivity*. We define the following role-values:

- *prov-said:Influencer* denotes the role of an agent that was used by an *prov-said:InfluenceActivity* that was associated with another agent. This means that the first agent influences the latter.

- *prov-said:Influencee* denotes the role of an agent that was associated with an *prov-said:InfluenceActivity*. This agent is being influenced by another agent used by the same *prov-said:InfluenceActivity*.

Following that, we define two subtypes of *prov-said:Influencee* and *prov-said:Influencer* respectively. Firstly, we model the follow relationship with the roles: *Follower* and *Followee* and secondly we model the activity of interaction by exchanging messages with the roles: *InteractionInfluencee* and *InteractionInfluencer*. Note that these roles are pairwise complementary by revealing the active behaviour of one agent in order to establish connections and to forward messages (follower, interactionInfluencee) and the passive behaviour of another (followee, interactionInfluencer) who exerts some influence on the first.

- *prov-said:Followee* denotes the role of an agent that was used by a *prov-said:FollowActivity* associated with another agent. This means that the latter followed the first.
- *prov-said:Follower* denotes the role of an agent that was associated with an *prov-said:FollowActivity*. This means than an agent establishes a unidirectional connection with another agent in social media.
- *prov-said:InteractionInfluencer* denotes the role of an agent that was used by an *provsaid:InteractionInfluenceActivity* associated with another agent. This means that the first agent is influencing the latter so that the latter propagates the messages of the first.
- *prov-said:InteractionInfluencee* denotes the role of an agent that was associated with an *prov-said:InteractionInfluenceActivity*. This means that the agent is being influenced by another agent by forwarding the messages of the latter.

We demonstrate these roles in Example 6.

*Example 6: Influence Roles.*

```
used(alice-follows-carol, twitter:Carol,
   [prov:role='prov-said:Followee'])
wasAssociatedWith(alice-follows-carol,
   twitter:Alice,
   [prov:role='prov-said:Follower'])

used(bob-influencedby-alice,
   twitter:Alice,
   [prov:role='prov-said:InteractionInfluencer'])
wasAssociatedWith
   (bob-influencedby-alice,
   twitter:Bob,
   [prov:role='prov-said:InteractionInfluencee'])
```

At this point we express the following constraints:

- We ensure that an *prov-said:InfluenceRelationship* always implies a *prov-said:InfluenceActivity*, *prov:Usage* and *prov:Association*. According to the type of *prov-said:InfluenceActivity*, specific *prov:roles* are being used.
- A *prov-said:InteractionInfluenceActivity* starts (and ends, since it is defined to be instantaneous) with the emission of *prov-said:CopiedMessage* or *prov-said:RevisedMessage*.

With these concepts, we have covered the model of influence with its possible expressions in activities, relationships and roles.

## 6. CONCLUSION AND FUTURE WORK

To sum up, PROV-SAID enables systems that analyze social media to incorporate provenance data in their information diffusion analysis. This will benefit the massive human-centric efforts for judging relevance and trustworthiness of information by unraveling its sources and intermediate steps.

For future work, we plan to investigate the necessary concepts to support broader cases of interaction among users (such as replies and mentions). This could allow to address more use cases, such as Q&A forums. We also plan for extensive evaluations with realistic social media data in order to validate our proof-of-concept, described in this paper.

## 7. REFERENCES

[1] F. Abel, C. Hauff, G.-J. Houben, R. Stronkman, and K. Tao. Semantics+ filtering+ search= twitcident. Exploring information in social web streams. In *HT*, pages 285–294, 2012.

[2] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts. Everyone's an influencer: quantifying influence on twitter. In *WSDM*, pages 65–74, 2011.

[3] R. A. Baños, J. Borge-Holthoefer, and Y. Moreno. The role of hidden influentials in the diffusion of online information cascades. *EPJ Data Science*, 2(1):1–16, 2013.

[4] G. Barbier, Z. Feng, P. Gundecha, and H. Liu. Provenance data in social media. *Synthesis Lectures on Data Mining and Knowledge Discovery*, 4(1):1–84, 2013.

[5] J. Cheney, L. Chiticariu, and W.-C. Tan. *Provenance in databases: Why, how, and where*, volume 4. Now Publishers Inc, 2009.

[6] J. Cheney and W3C Provenance Working Group. Semantics of the PROV Data Model. W3C Note 30 April, 2013.

[7] P. Cheney, J. Missier, Moreau, L.(Eds.), and W3C Provenance Working Group. PROV-CONSTRAINTS: Constraints of the PROV Data Model. W3C, 2013.

[8] T. De Nies, S. Coppens, E. Mannens, and R. Van de Walle. Modeling uncertain provenance and provenance of uncertainty in w3c prov. In *WWW (Companion Volume)*, pages 167–168, 2013.

[9] B. Glavic, K. Sheykh Esmaili, P. M. Fischer, and N. Tatbul. Ariadne: Managing fine-grained provenance on data streams. In *Distributed event-based systems*, pages 39–50, 2013.

[10] A. Guille, H. Hacid, C. Favre, and D. A. Zighed. Information diffusion in online social networks: A survey. *SIGMOD Rec.*, 42(2), 2013.

[11] T. Huynh, P. Groth, S. Zednik(Eds.), and W3C Provenance Working Group. PROV Implementation Report. W3C Working Group Note 30 April, 2013.

[12] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *SIGKDD*, pages 137–146, 2003.

[13] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang. Prominent features of rumor propagation in online social media. In *ICDM*, pages 1103–1108, 2013.

[14] S. Magliacane, P. T. Groth, et al. Towards reconstructing the provenance of clinical guidelines. In *SWAT4LS*, 2012.

[15] C. Maumet, G. Flandin, B. Nichols, J. Steffener, K. Helmer, et al. Extending NI-DM to share the results and provenance of a neuroimaging study: Implementation within SPM and FSL. *Frontiers in Neuroinformatics*, 2014.

[16] P. Missier and Z. Chen. Extracting PROV provenance traces from wikipedia history pages. In *EDBT/ICDT (Workshops)*, pages 327–330, 2013.

[17] L. Moreau, P. Missier (Eds.), and W3C Provenance Working Group. PROV-DM: The PROV Data Model. W3C, 2013.

[18] S. A. Myers, C. Zhu, and J. Leskovec. Information diffusion and external influence in networks. In *SIGKDD*, pages 33–41, 2012.

[19] I. Taxidou and P. M. Fischer. Online analysis of information diffusion in twitter. In *WWW (Companion Volume)*, pages 1313–1318, 2014.

[20] I. Taxidou and P. M. Fischer. Rapid: A system for real-time analysis of information diffusion in twitter. In *CIKM*, pages 2060–2062, 2014.

[21] C. Wilson, A. Sala, K. P. Puttaswamy, and B. Y. Zhao. Beyond social graphs: User interactions in online social networks and their implications. *ACM Transactions on the Web*, 6(4):17, 2012.