Author Copy for personal use. For final version, please check: doi: 10.1109/SCOPE-GCTC.2018.00012
H. Purohit, S. Nannapaneni, A. Dubey, P. Karuna and G. Biswas, "Structured Summarization of Social Web for Smart Emergency Services by Uncertain Concept Graph," 2018 IEEE International Science of Smart City Operations and Platforms Engineering in Partnership with Global City Teams Challenge (SCOPE-GCTC), Porto, Portugal, 2018, pp. 30-35.

Structured Summarization of Social Web for Smart Emergency Services by Uncertain Concept Graph

Hemant Purohit†, Saideep Nannapaneni*, Abhishek Dubey‡, Prakruthi Karuna†, Gautam Biswas‡

†Department of Information Sciences and Technology, George Mason University, Fairfax, VA, USA
*Department of Industrial, Systems and Manufacturing Engineering, Wichita State University, Wichita, KS, USA
‡Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, USA
†{hpurohit,pkaruna}@gmu.edu, *saideep.nannapaneni@wichita.edu,
‡{abhishek.dubey,gautam.biswas}@vanderbilt.edu

***Abstract*—The Web has empowered emergency services to enhance operations by collecting real-time information about incidents from diverse data sources such as social media. However, the high volume of unstructured data from the heterogeneous sources with varying degrees of veracity challenges the timely extraction and integration of relevant information to summarize the current situation. Existing work on event detection and summarization on social media relates to this challenge of timely extraction of information during an evolving event. However, it is limited in both integrating incomplete information from diverse sources and using the integrated information to dynamically infer knowledge representation of the situation that captures optimal actions (e.g., allocate available finite ambulances to incident regions). In this paper, we present a novel concept of an Uncertain Concept Graph (UCG) that is capable of representing dynamic knowledge of a disaster event from heterogeneous data sources, particularly for the regions of interest, and resources/services required. The information sources, incident regions, and resources (e.g., ambulances) are represented as nodes in UCG, while the edges represent the weighted relationships between these nodes. We then propose a solution for probabilistic edge inference between nodes in UCG. We model a novel optimization problem for the edge assignment between a service resource to a region node over time trajectory. The output of such structured summarization over time can be valuable for modeling event dynamics in the real world beyond emergency management, across different smart city operations such as transportation.**

***Index Terms*—Social Media, Event Summarization, Disaster, Resource Allocation, Uncertain Concept Graph**

I. Background and Motivation

The growth of the Web and social media platforms have opened new opportunities for understanding the world events [1], [2]. Especially during disasters, the recent years have observed a large-scale, unprecedented sharing of data for situational updates via social media beyond the conventional official reporting channels and news media [3], [4]. We consider large-scale disaster events (natural or man-made) such as hurricanes, cyclones, and terrorist activities in this research. During such events, the number of people calling for help via traditional help lines such as 911 in U.S. often becomes unmanageable by the human resources of response agencies (e.g., as reported during the hurricane Harvey disaster in 2017 [1]). The limitation of traditional communication lines has resulted in people resorting to alternative channels for seeking help, providing situational updates, and also responding to needs for help [5], [6], [7].

Response agencies, therefore, have started exploring ways to leverage social media for enhanced situational awareness and decision support during a disaster [8]. For instance, given the available resources (e.g., ambulances) are limited during a disaster response, the agencies expect to improve their resource dispatch mechanisms while considering all available information via both traditional helplines like 911 and novel social media.

A key challenge for social media mining to enhance operational decision support of the response agencies is to efficiently infer relevant knowledge of the dynamic situation. Solutions for extraction and integration of information from heterogeneous data streams, particularly event detection and summarization methods [9], [10], [11] are useful to address this challenge. However, existing research on event summarization is limited to create timely and novel unstructured text summaries primarily for human readers and lacks a focus on structured textual summaries with uncertainties in mining the structured information.

Dynamic Disaster Context: It is important to understand a disaster situation before describing a representation to summarize its evolving knowledge. Consider three information sources I_1, I_2 and I_3, who post messages on Twitter for some incidents E_1 and E_2 with location (latitude-longitude) metadata in a region R_1 (assume affected area is partitioned into 4 regions R_1, R_2, R_3 and R_4). The goal of response agencies is to allocate relevant resources (e.g., ambulances for medical type incidents) to the region R_q where an incident has occurred. While we can specifically map a message of I_k to a region R_q by determining which region's bounding box contains the location of the I_k's message, there is uncertainty in extracting information for the incident and the required resource from the natural language text of the message. Using text mining models[12], [4], we can predict the message received from a source I_k for: (a.) likelihood of an incident occurrence of type E_b, and (b.) likelihood of a required resource type S_a. Also, the veracity of the information source (e.g., Red Cross versus a common citizen account on Twitter) can be estimated using

[1]http://wapo.st/2iAZhbc?tid=ss_tw&utm_term=.fdd1e12b125a

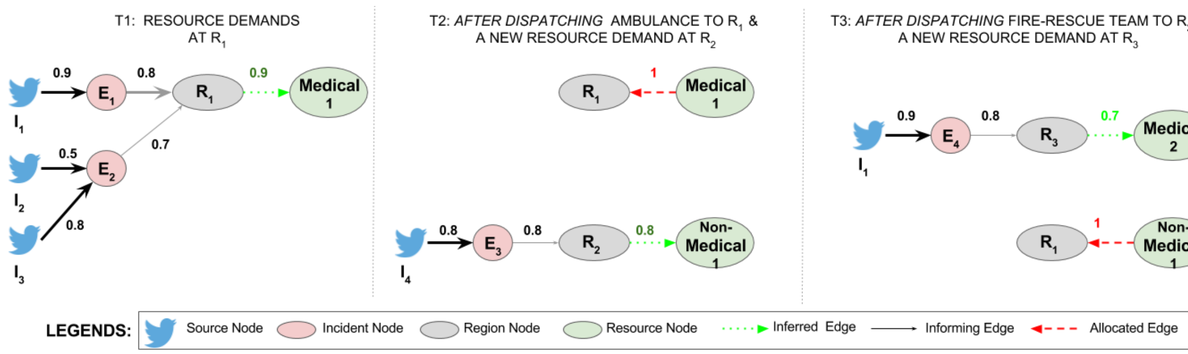

Fig. 1. An illustration of a disaster situation over time trajectory with unsteady nodes and edges. At a time instant T_x, information source node I_k shares a Twitter message about an occurrence of an incident E_b at a region R_q, extracted by text mining models with uncertainty, and leads to probabilistic (*Informing*) edges. The relation of a required resource S_a^i (medical or non-medical) to R_q at T_x is inferred by aggregating all observed incidents of a resource type need in R_q (depicted by *Inferred* edge) and is updated after the dispatch decision (depicted by *Allocated* edge).

credibility models (ibid) to provide weights to the extracted information as well. By aggregating the likelihood of incident occurrences and the required resource demands over all weighted information sources in a time interval, for instance T_1, the sources I_1, I_2 and I_3 are mapped to a region R_1; the response agencies then can get an expected likelihood for making a decision to dispatch a relevant resource (e.g., $Medical1$) after T_1.
Given the various uncertainty types in extracting information from streaming data (both the predicted likelihood of the required resource demands and the veracity degree of the information sources), there is a need to capture the spatio-temporally varying relationships among regions, the required resource types, and the information sources. Furthermore, there is a need to efficiently represent and summarize the updates in the relationships after resource allocations over time trajectory. Figure 1 illustrates this dynamic scenario.

Contributions: The overall contribution of this paper is the position of a novel framework for structured summarization of dynamic event contexts (with a focus on emergency services). The framework allows the fusion of information from heterogeneous sources while representing uncertainty in the extracted information from the high volume of sparse, redundant data of heterogeneous sources with varying veracity.

Specifically, the proposed approach for temporal structured summarization of heterogeneous data streams – *Uncertain Concept Graph* (UCG) captures knowledge representation of spatio-temporal relationships between information sources, resource requests (contingent on reports of incident types), spatial regions, and the available resources during disasters. UCG enables modeling of the uncertainty in information extraction from multimodal content, by mining natural language text and multimedia. We formulate a solution framework to sequentially update the resource requests and optimally dispatch resources to locations, reflected by updating the edge relationships in the UCG.

Paper Organization: The remainder of the paper is organized as follows. Section II discusses the methods and technical solutions for the problem of structured summarization of resource needs using the *UCG*. Section III formulates an optimization problem for resource allocation using the uncertain concept graph UCG. Conclusions and Future work follows in Section IV.

II. Methods and Technical Solutions

In this section, we first discuss the problem, limitations of existing methods, and then, describe the novel method of UCG representation.

A. Problem of Structured Summarization

Dynamic disaster contexts require the representation and integration of streaming information from content of heterogeneous sources such that it helps summarization of the evolving events in a structured form, which can feed into decision support systems over time. Such structured form of information needs to be extracted from semi-structured data streams of heterogeneous information sources, while taking into account the uncertainties in the information sources and extraction models.

B. Limitations of Existing Work

Prior research on event detection and summarization techniques [1], [9], [2], [13], [14], [15], [10], [16], [17], [18], [19] has focused primarily on efficient characterization of events for identification and textual summarization of events using clustering approaches for an online reader. We provide a brief overview of the diverse approaches. [2] presented an approach to identify evolutionary events from social streams by using topical content of messages and the graphical structure of the dynamic network of interactions among message authors. [16] presented an event identification as well as ranking method of burst events and localized events for a region and a time-frame. [19] recently proposed a model for including lack of activity or absenteeism instances for event detection in social media. [9] focused on detecting events on Twitter and extracting a few tweets that best summarize the chain of relevant occurrences. [13] proposed an unsupervised joint topic modeling method to summarize trending subject matter by jointly discovering the representative and complementary information from news and tweets. [14] proposed a framework to extract relevant representative tweets from an unfiltered tweet stream to generate a coherent and concise summary of an event. [15] proposed a summarization model for news content to provide answers to readers for 'what, when, why, where, how' questions of events. A recent survey by [10] summarizes the challenges of analyzing evolutionary network representation with online information streams, under the one-pass constraint of data streams.

Dynamic Bayesian Network (DBN) [20] presents an alternative approach to model temporal dynamics of a system for dynamic disaster context, however due to the unsteady nodes and edges for representing the system state as well as the lack of apriori knowledge for the incidents and information sources for a region, it is not feasible to adapt a DBN. For instance, the volunteers providing information for incidents in a region can appear and disappear with time and the probability distributions associated with the links are not stationary. Therefore, the DBN assumptions are violated. Complementing prior research on evolutionary event context, we present a model for structured summarization in disasters.

C. Proposed Solution: Uncertain Concept Graph (UCG)

We describe below the representation of various uncertainties corresponding to a disaster context by constructing a novel graphical model known as uncertain concept graph *UCG* and present the characteristics of several nodes and edges in it.

Definition 1: An Uncertain Concept Graph (UCG) is a time variant probabilistic graphical model defined over a set of nodes $\mathbb{V}$. The UCG model at a given time instant T_m is written as UCG_{T_m}=(V_m, L_m, P_m), where $V_m \subseteq \mathbb{V}$ is the set of nodes and the edge or link set $L_m \subseteq V_m \times V_m$ describes the directional conditional relationship between the nodes. The probability distribution $P_m^e \in P_m$, $e \in L_m$ describes the relative probabilistic weight of a target node connected to the edge e with respect to a source node connected to the edge e. An external observation process and an internal update process control the evolution of the UCG across time. During this update the nodes and the edges of the UCG can change.

For simplicity we assume that UCG evolves sequentially with either the observation step or the update step at a time instant T_m. The node set $\mathbb{V}$ is flexible to represent different categories of concepts from domain knowledge. The node set V represents concept nodes of *Resources*, *Regions*, *Incidents*, and *Information Sources* as illustrated in Figure 1. The edge set represents the relationships between information sources and incident nodes as well as incident nodes and region nodes as *Informing edges*. The relationship between the region nodes and the resource nodes is represented as *Inferring edges* as well as *Allocation edges* if resource allocation is confirmed.

In the following, we first formally define variables for representing information in UCG, characteristics of nodes and edges of UCG, and then, propose an optimization problem for inferring relationships between the *Region* and *Resource* nodes at a time instant T_m in order to help summarize resource needs at T_m.

- Incident Node (E_b): A node that represents an undesirable event that occurred in a region.
- Region Node (R_q): We assume that the entire community is divided into several non-intersecting areas. Each non-intersecting area is referred to as a Region.
- Service Resource Node (S_a): A resource for response services (such as medical or fire-rescue) that is desired in responding to an incident in a region. We consider medical and non-medical resource types for simple explanation.
- Information Source Node (I_k): Any entity that provides information regarding an incident. For instance, sources can provide *tweets* from Twitter, *images* from Instagram, *posts* from Facebook, and 911 reports.
- Source Confidence: The degree of confidence that is assigned to a source. For example, 911 operators and emergency services' accounts on Twitter can be associated with higher confidence than a random information source.
- Service Time: The time spent by the emergency service personnel to mitigate the damage due to an incident in a region.
- Travel Time: The time that the emergency services take to reach the destination from the current location.
- Incident Region: The region where an event occurred. Note that there could be several incident locations in the affected community.
- Auxiliary Region: The regions that contain the central locations of the emergency services such as fire stations and hospitals.

D. UCG_{T_m} Nodes

From any information source in a time interval T_m (denoted by the last time instant of the interval), we extract four types of details from its message (e.g., *tweet*) to construct nodes and edges of UCG_{T_m}: (1) the location of informing source, (2) the time when the information was received, (3) the incident type, and (4) the type of service resource need (e.g., medical).

Due to high volume of social media sources during disasters, we require machine learning models to automatically extract the incident details for dynamic situation awareness. However, when such techniques are used, there exists uncertainty in the model predictions. Furthermore, it is possible to have varying degrees of veracity with information sources on a social media (e.g., based on user affiliations and influence.) On the other hand, the traditional helplines such as 911 calls enable relatively accurate sourcing of information for the situation details of a region due to the direct communication with the 911 operators. Given the flexible representation in UCG, such heterogeneous information sources can be easily modeled as information source nodes with different veracity.

Let $I_k, k = 1, 2..N$ represent N distinct information sources, such as N Twitter users who share any tweet for an incident in a region. For the service resource nodes, let $S_a, a = 1, 2, ..A$ represent A types of resources available. For simplicity, we assume $A = 2$ and the resource types are medical (e.g., ambulances) as S_1 and non-medical (e.g., fire trucks) as S_2. In this work, we divide the entire community into Q regions, denoted as $R_q, q = 1, 2, ..Q$, given that the usual operations of response agencies are driven by predefined jurisdictions. Also, as the exact time when the incident occurred in a region R_q is not known without on-site observation, we assume it as the time when the first information source for requesting a specific service type in R_q is available (such as the time of first tweet for the incident).

E. UCG_{T_m} Edges

An edge in UCG is the most significant contributor for summarizing a situation at a region R_q at a given time T_m. To model the edge relationships between UCG nodes, we rely on the automated text mining models to get prediction probabilities of information extraction, in order to create and weigh the edges from a source node to incident nodes. We propose a measure for edge relationship between a region node R_q and a resource type S_a based on the incident reports sourced from information sources I_k as:

$$P(R_q, E_b, S_a) = \Sigma_k P(R_q, E_b, S_a|I_k) \times P(I_k) \quad (1)$$

where $P(R_q, E_b, S_a|I_k)$ represents the joint probability that an information source I_k is requesting a resource S_a in a region R_q for an incident type $E_b, b = 1, 2, .., B$. Note that the uncertainty in the resource and location details is due to the model prediction and not from the information source itself. Therefore, the overall probability of a resource requirement in a given region can be obtained by aggregating all available information from diverse source nodes, with the prior for an information source as a measure of veracity of the source.

In this way, the probability of resource requirements at every region R_q can be computed. Here, $P(I_k)$ refers to the probability mass function of a weight I_k, which is equal to the degree of confidence in that information source. Let $T = T_m^{init}$ represents the time of the first incident at several regions. Note that it is possible an incident may not have happened at some regions (hence, the spatial-temporal variation in a state of UCG_{T_m}.) It is assumed that there exists new information sources associated with any new incidents at the same as well as different regions over future time intervals. In such cases, the probability $P(R_q, E_b, S_a)$ would increase due to the increased requirement for resources in the future if no resource allocation is done based on UCG_{T_m} state. We also make a Markov assumption to carry over this likelihood of resource demand to the future time intervals (T_{m+x}, $x > 0$).

F. Matrix Data Structure for UCG_{T_m}

To capture the dynamic relationships in UCG via edges and probabilistic edge weights, we propose matrix representation of UCG. For computational tractability, the matrix form assists in representing UCG nodes and edges over time trajectory.

1) $[M1]$ *Region - Resource Matrix*: It represents the assignment state of a resource node to a region node in UCG, and can have an entry as 0 or 1. It is the most important matrix for resource need summarization at T_m. At the initial time over trajectory, all the resources are assumed to be present in auxiliary regions, and when incidents occur, they are sent to the incident regions.
2) $[M2]$ *Region - Incident Matrix*: It represents the probability that an incident type is occurring at a location in a region. Each entry is a value between 0 and 1.
3) $[M3]$ *Region - Resource Time Matrix*: It represents the time it would take for a resource to provide services at a region. Therefore, the time is equal to the summation of: (1) remaining service time at the current region, and (2) travel time to the new region. Service times may vary depending on the resources. Some resources can be equipped with advanced facilities compared to others. The rows represent service resources whereas the columns represent incident locations.

When a resource is deployed, it becomes unavailable for a period of time until it finishes the current assignment task (allocation state captured by matrix $M1$), which in turn, affects the availability of resources at any given time interval T_m. Therefore, real-time region to resource allocation analysis, i.e. inference of values in matrix $M1$, is required in the presence of uncertainty in information represented in UCG_{T_m} and the time-varying resource availability, which leads to an optimization problem. We address this challenge via an inference problem formulation in the following.

III. UCG_{T_m} Node-Edge Updates: Optimization Problem

Our goal is to infer the matrix $M1$ at any given time, which affects the potential service allocation and delays leading to possible damages from an incident. When disasters occur, they result in several types of damage such as loss of human lives and property losses. We formally divide the damage into two types after a disaster occurrence - avoidable damage (D_A) and unavoidable damage (D_U). Avoidable damage refers to the losses that occur due to the unavailability of emergency resources whereas unavoidable damage refers to the losses due to the disaster that cannot be controlled. The total damage D can be computed as the summation of avoidable and unavoidable damage. When all the required emergency services are available at the time of disaster, then the total damage is equal to the unavoidable damage.

In this work, we assume that the avoidable damage D_{A_q} at a region R_q increases with the time lag between first incident reporting and corresponding service allocation at an exponential rate, and define it as:

$$D_{A_q}(\Delta T_q) = \alpha \exp(-\lambda \Delta T_q) \quad (2)$$

where α and λ represent the model parameters that are assumed available from historical records or previous disaster reports, and the number of information sources. If T represents the current time and T_q^{init} the first incident reporting time, then $\Delta T_q = T - T_q^{init}$.

As new information sources are available with time, these parameters will be updated based on any new information. It can be assumed that the severity of the disaster in a region is proportional to the number of distinct information sources.

The exponential function used here is similar to that used in survival analysis [21].

We consider the two service resource types (S_a, a=1,2) in this work as ambulances (S_1) for medical-related incidents and fire trucks (S_2) for non-medical related incident needs. Let f and g represent the number of fire trucks and ambulances available for service. The individual fire trucks and ambulances are represented as $S_1^i, i = 1, 2...f$ and $S_2^j, j = 1, 2...g$ respectively.

The resources are required for dispatching to the regions wherever their services are the most needed at time T_m. This is achieved by solving the following optimization problem for the minimum utility as:

$$\text{Minimize} \\ E\Big[\Big(\sum_a \sum_i \sum_b \sum_q w_{S_a^i} \times P(R_q, E_b, S_a^i) \times D_{A_q}(\Delta T_q)\Big) + \Big(\sum_a \sum_i (1 - w_{S_a^i}) \times C(S_a^i, t_1^i)\Big)\Big] \tag{3}$$

In the above optimization formulation, $w_{S_a^i}$ is a penalty weight factor (0 or 1) for a resource type's allocation decision. Penalty factor is 1 for not allocating a resource S_a^i to a region R_q in which case, the expected demand likelihood (first part of the utility objective) is a critical part contributing to the overall cost. ΔT_q refers to the time lag from the disaster incident reporting to the service resource arrival in the region R_q. The goal is to mobilize all available resources such that the overall avoidable damage and the overall costs are minimized. $C(.)$ represents the cost associated with a resource movement. It takes two inputs - the resource with its current location and the estimated time a resource might take to reach and provide services at a particular region. This time is equal to the summation of remaining service time at the current location, travel time to the new region, and the estimated service time at a new region. Also, the incident severity level in the emergency domain (such as Level 1 or Level 2 or Level 3, b=1,2,3) at a region is often not known precisely and thus, each level can be associated with a probability. To estimate the overall avoidable damage, we sum over all incident levels.

Due to the presence of uncertainty in the travel and service times, the overall cost function is also uncertain. Therefore, we minimize the expectation value of the overall cost in the proposed optimization objective (equation 3).

Optimization solution: We perform the optimization analysis at specific time instants (or the end of time intervals) for inferring values of matrix $M1$. At the end of time interval T_m, we update the matrices $M2$ and $M3$ for service and travel times based on distance measures as well as incident likelihood with resource types needed based on real-time information sources. The control variables in the optimization formulation are the discrete values which represent if a resource is dispatched to a particular region. The dispatch of a resource to a particular region depends on the cost associated with the time it requires to reach and serve at a particular region from the current time T. We assume that real-time location tracking of the emergency service resources is available to the response agencies. Therefore, after providing the necessary assistance at a particular location, the emergency service resources can be routed to a different location from the previous location without having to come back to the central station in the auxiliary region. However, in some cases such as ambulances, they cannot go directly from one incident region to another region, but may have to return to the hospitals to assist people that require medical attention. Thus, the travel time from the auxiliary region of a resource to an incident region includes the travel time from a previous region to the hospital and from the hospital to the new region.

As mentioned above, the optimization analysis will be carried out sequentially at subsequent time intervals (which can be equally spaced). Let $T_m, m = 1, 2...U$ represent the U time instants when the optimization analysis is carried out. Therefore, $T_m - T_{m-1} = T_{m+1} - T_m$ and we also refer a time interval by its last time instant, e.g., T_m and T_{m+1} in the previous example. As the probability of resource requirement in a region ($P(R_q, E_b, S_a)$) changes with time, we represent that probability at any given time as $P(R_q, E_b, S_a)_T$, where the subscript represents real time. At any last time instant of a time interval T_m, when the optimization is carried out, we calculate the $P(R_q, E_b, S_a)_{T_m}$ using the probability at the previous time step when the optimization was carried out, i.e., T_{m-1} as well as all the new information that was obtained in the time interval, $\{T_{m-1}, T_m\}$. Thus, the optimization analysis results into UCG_{T_m} state that represents the structured summarization of the dynamic disaster context over time.

IV. Conclusions and Future work

This paper is a position statement paper where we reviewed the existing literature regarding summarization analysis of social web and identified the potential issues with respect to using the existing methods in emergency response context. We presented a novel approach of Uncertain Concept Graph (UCG) to summarize heterogeneous data streams during a disaster event, by efficiently representing the knowledge of dynamic situation and inferring probabilistic relationships between the key concepts of regions, resources, and sources. We modeled a UCG via a probabilistic graph framework and formulated a service/resource need and allocation at a region on a given time trajectory using a novel optimization problem formulation. An immediate future work of interest is the demonstration of the proposed framework for real-world disaster events. To this end, we will experiment with the social media and Web data collected during the events of recent hurricane season of 2017.

V. Acknowledgement

The research reported in this paper is supported by funds from the National Science Foundation (NSF) under the award number IIS-1657379 and CNS-1640624 as well as the Vanderbilt Initiative for Smart-City Operations Research (VISOR).

References

[1] Q. Zhao, P. Mitra, and B. Chen, "Temporal and information flow based event detection from social text streams," in *AAAI*, vol. 7, 2007, pp. 1501–1506.

[2] C. C. Aggarwal and K. Subbian, "Event detection in social streams," in *Proceedings of the 2012 SIAM international conference on data mining*. SIAM, 2012, pp. 624–635.

[3] L. Palen and K. M. Anderson, "Crisis informatics - new data for extraordinary times," *Science*, vol. 353, no. 6296, pp. 224–225, 2016.

[4] C. Castillo, *Big Crisis Data: Social Media in Disasters and Time-Critical Situations*. Cambridge University Press, 2016.

[5] J. Yin, A. Lampert, M. Cameron, B. Robinson, and R. Power, "Using social media to enhance emergency situation awareness," *IEEE Intelligent Systems*, vol. 27, no. 6, pp. 52–59, 2012.

[6] H. Purohit, C. Castillo, F. Diaz, A. Sheth, and P. Meier, "Emergency-relief coordination on social media: Automatically matching resource requests and offers," *First Monday*, vol. 19, no. 1, 2013.

[7] M. Imran, C. Castillo, F. Diaz, and S. Vieweg, "Processing social media messages in mass emergency: A survey," *ACM Computing Surveys (CSUR)*, vol. 47, no. 4, p. 67, 2015.

[8] S. U.S. Homeland Security Science & Technology, "Using social media for enhanced situational awareness and decision support," 2014, accessed: 2017-06-12.

[9] D. Chakrabarti and K. Punera, "Event summarization using tweets," *ICWSM*, vol. 11, pp. 66–73, 2011.

[10] C. Aggarwal and K. Subbian, "Evolutionary network analysis: A survey," *ACM Computing Surveys (CSUR)*, vol. 47, no. 1, p. 10, 2014.

[11] X. Chang, Y. Yang, G. Long, C. Zhang, and A. G. Hauptmann, "Dynamic concept composition for zero-example event detection," in *AAAI*, 2016, pp. 3464–3470.

[12] H. Purohit, G. Dong, V. Shalin, K. Thirunarayan, and A. Sheth, "Intent classification of short-text on social media," in *Smart City/SocialCom/SustainCom (SmartCity), 2015 IEEE International Conference on*. IEEE, 2015, pp. 222–228.

[13] W. Gao, P. Li, and K. Darwish, "Joint topic modeling for event summarization across news and social media streams," in *Proceedings of the 21st ACM international conference on Information and knowledge management*. ACM, 2012, pp. 1173–1182.

[14] F. C. T. Chua and S. Asur, "Automatic summarization of events from social media," in *ICWSM*, 2013.

[15] G. Binh Tran, "Structured summarization for news events," in *Proceedings of the 22nd International Conference on World Wide Web*. ACM, 2013, pp. 343–348.

[16] W. Feng, C. Zhang, W. Zhang, J. Han, J. Wang, C. Aggarwal, and J. Huang, "Streamcube: hierarchical spatio-temporal hashtag clustering for event exploration over the twitter stream," in *Data Engineering (ICDE), 2015 IEEE 31st International Conference on*. IEEE, 2015, pp. 1561–1572.

[17] Z. Li, J. Tang, X. Wang, J. Liu, and H. Lu, "Multimedia news summarization in search," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 7, no. 3, p. 33, 2016.

[18] S. Ghosh, P. Chakraborty, E. O. Nsoesie, E. Cohn, S. R. Mekaru, J. S. Brownstein, and N. Ramakrishnan, "Temporal topic modeling to assess associations between news trends and infectious disease outbreaks," *Scientific reports*, vol. 7, p. 40841, 2017.

[19] F. Jin, F. Chen, R. P. Khandpur, C.-T. Lu, and N. Ramakrishnan, "Absenteeism detection in social media," in *Proceedings of the 2017 SIAM International Conference on Data Mining*. SIAM, 2017, pp. 606–614.

[20] K. P. Murphy, "Dynamic bayesian networks," *Probabilistic Graphical Models, M. Jordan*, vol. 7, 2002.

[21] R. G. Miller Jr, *Survival analysis*. John Wiley & Sons, 2011, vol. 66.