First Workshop on Intelligent Transportation Informatics at the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR) 2018.

Modeling Transportation Uncertainty in Matching Help Seekers and Suppliers during Disasters

Hemant Purohit
George Mason University
hpurohit@gmu.edu

Nikhita Vedula
Ohio State University
vedula.5@osu.edu

Krishnaprasad Thirunarayan
Wright State University
tkprasad@knoesis.org

Srinivasan Parthasarathy
Ohio State University
srini@cse.ohio-state.edu

ABSTRACT

Social web empowers citizens to express urgent calls for needs and offers of help during disasters. Identifying and matching those who need help with those who offer help for the required resources in real time on the social web has multiple outcomes. Such a matching can assist responding agencies in realistic estimation of needs and allocation of resources by region. It can also enable individual users in making decisions of where and whom to offer help, thus, supporting overall 'citizen-driven' relief coordination (e.g., #OccupySandy, #harveyrelief on Twitter). However, it is challenging to timely match users online and connect them offline due to the incomplete information and dynamic unavailability of transportation routes, that result from the damage of an evolving disaster. In this paper, we address these challenges by formulating a novel problem of multi-objective bipartite matching, and proposing a generalizable solution of a Social Matching Framework (SMF). The SMF model formalizes different sources of uncertainty in a bipartite network representation, which consists of nodes denoting users with complementary intents of *seeking help-offering help* and edges denoting transportation paths between user locations. An edge weight models the cost of traveling the distance of a path and the uncertainty of its existence during a disaster. Our approach provides pareto-efficient path recommendations for connecting help-offering to help-seeking users via minimum distance distance and minimum uncertainty paths, by pre-computing and parallelizing the steps in updating edge weight attributes for transportation cost and path availability. We demonstrate the SMF applicability for the real world using Twitter data, and show that such a match recommendation system can greatly reduce the efforts of volunteer teams and response organizations.

CCS CONCEPTS

• **Information systems → Social networks**; *Novelty in information retrieval*; *Enterprise applications*; • **Applied computing →** E-government;

KEYWORDS

Social Media, Disaster Relief, Help Intent, Bipartite Matching

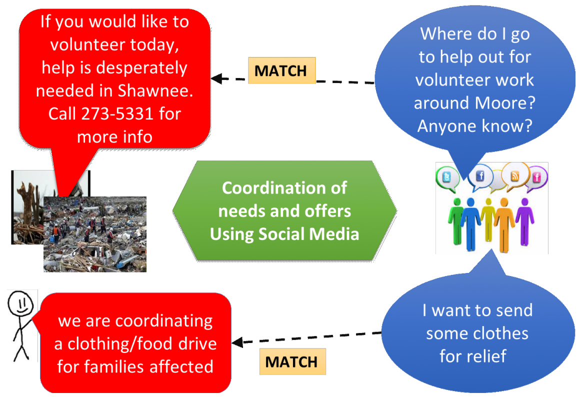

Figure 1: Matching of complementary user intent pairs

1 INTRODUCTION & BACKGROUND

Social Web facilitates a platform for the public to self-organize, as shown during mass events, including crises and social movements [7]. During natural disasters, one aspect of such self-organizing behavior includes volunteer-driven relief efforts towards awareness and connectivity of those who need help and those who offer help as observed in recent disasters (e.g., Twitter hashtags #OccupySandy [1] and #JKFloodRelief [6]). For instance, people who are trying to help the affected population in a disaster by offering food rations could be connected to those who are seeking help for food, and thus, achieve better disaster relief objectives while reducing burden on the formal or governmental response agencies [5, 8] (see Figure 1). Similarly, the situational awareness of the resultant resource needs and availabilities can help prevent the "second disaster" of unsolicited resource management (as observed for surplus clothing donations after Hurricane Sandy [1]).

Organized-relief efforts [1] of volunteer groups (e.g., digital humanitarian network [3]) in recent years have provided a direction to use online media to help coordinate relief by matching user needs and offers via manual, crowdsourcing approaches. However, such initiatives cannot scale due to the uncertainty and dynamism of the emerging needs in different locations, and the availability of volunteers offline and online with resources. Portals such as *Recovers.org* provide a mechanism for self organizing communities to facilitate a community-led disaster relief efforts for resource management. However, such relief efforts also cannot scale to their full potential due to missing and uncertain information for transportation networks, as well as limited human resources for manual matching in near real time. Hence, automatically matching the users on social and web platforms with complementary help intents can prove to be extremely helpful [5, 8]. However, this is challenging due to uncertainty in determining the intent and semantics of correct resource matching, and also, practical constraints on the transportation feasibility of matching for efficient disaster relief. For instance, matching a food drive in New York city to a person seeking food

[1]http://www.npr.org/2013/01/09/168946170/thanks-but-no-thanks-when-post-disaster-donations-overwhelm

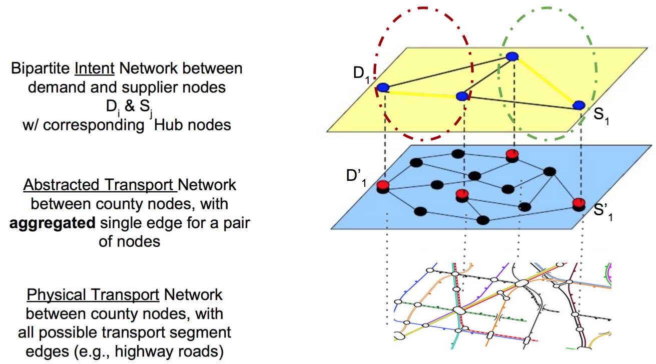

Figure 2: Representing a virtual bipartite intent network by an abstraction of the physical transport network

help in Haiti is not helpful for near real time relief coordination. Thus, while there has been extensive research on the matching problem in computing and operations research disciplines, the problem being addressed here has unique characteristics of uncertainty and matching constraints not yet dealt with in the literature.

The main contribution of this paper is to formulate a novel problem of linking help *seeking-offering* users on social media using a multi-objective bipartite matching approach, and propose a generalizable solution of a Social Matching Framework (SMF). The SMF model formalizes different sources of uncertainty constraints during disaster relief, in particular the availability of offline transportation network via the edge weight updates in the bipartite network. Our solution provides pareto efficient path recommendations for connecting help-*offering* to help-*seeking* users with minimum distance and minimum uncertainty in path existence. We now describe our overall SMF model, followed by its experimental validation.

2 SOCIAL MATCHING FRAMEWORK

We propose a quantitative solution through a novel social matching framework to address the following problem:

Problem Statement: Given a bipartite network $G = (D, S, E)$ with a transportation cost matrix TC of dimension $|D| \times |S|$ based on a cost function $C(e) \mid e = (d, s)$ and $e \in E$, a transportation network $N = (I, E')$ of nodes $i \in I$ with edges based on adjacency $e' \in E'$ for transportation paths, find a feasible matching $M_x \mid x \in [1, k]$ corresponding to each of the k resource categories (e.g., food, clothing, voluntary services) such that:
a.) $M_x \subseteq E$,
b.) the total transportation distance cost of M_x is minimum, and
c.) the total likelihood of transportation path existence is maximum.

This problem formulation is a multi-objective bipartite matching optimization. We propose a novel solution to address uncertainty in this problem by exploiting prior knowledge of transportation networks to augment a social media-driven bipartite solution. We propose to abstract the social media bipartite intent network and the physical transportation network into an abstract hub network (see Figure 2). Our framework incorporates four major components:

(1) *bipartite intent network creation* with user nodes having *seeking* or *offering* intent, and the node abstraction to the nearest county as hub nodes (inspired by the practical relief efforts),
(2) *transportation cost matrix creation* for possible transportation paths between a pair of matchable demand and supply nodes,
(3) *path connectivity matrix creation for estimating uncertainty* in connecting a pair of nodes, given the likelihood of damage for various roads or paths in the transportation network, and
(4) *local pairwise path selection and global matching optimization* for path recommendations using a pareto optimal solution.

2.1 Bipartite Intent Network Creation

We first create a bipartite intent network $G = (D, S, E)$ based on the identified nodes from social media, with *seeking* (demander) or *offering* (supplier) intent for a resource category such as clothing [5]. We extract the set of attributes for each user node in the bipartite network using metadata of messages with *seeking*/*offering* intent, such as location. We construct edges between a pair of nodes from the two bipartites based on similar resource category. We argue that it is inefficient for an individual supplier alone to travel to a demander. Thus, we propose to abstract an individual user to a 'relief hub' node, as shown in the second layer of Figure 2, for the purpose of aggregating the needs and offers at the nearest city or county center for an individual. Such a hub can be an NGO's or volunteer group's distribution and collection center, e.g., Red Cross centers. The intuition behind such an abstraction is two-fold: first, the scalable relief efforts are often supported by pre-established or temporary collection and distribution centers of the responding groups (e.g., during J&K floods an NGO Goonj had setup city-wide collection centers and distribution centers in the affected region [6]. Second, it is efficient to aggregate resource needs and offerings at hubs where local needs can be directly matched, to avoid individual level coordination efforts, which often lead to the duplication and waste of goodwill.

For such an abstracted bipartite network of 'hub' nodes, we assume the cost of transportation to and from an individual user node to the hub nodes as insignificant and can be discarded as desired. Lastly, we assume an abstracted hub node to inherit expected mean distribution of resource category needs from the related individual user nodes. We also consider only those hubs which demand or supply non-monetary resources.

Our solution builds on the abstracted transport network of hub (county) nodes (see Figure 2), utilizing a physical transportation network (e.g., road, rail, air) across county nodes in a nation. It helps estimate the transportation cost between two hub nodes, and the uncertainty in the connectivity between the nodes in the network.

2.2 Transportation Cost Matrix Creation

We define a cost matrix W of dimension $|D| \times |S|$ for the transportation cost between a matchable pair of users in the bipartite intent network as the cost of transporting relief supplies between their nearest hubs in the abstracted transport network. In this paper, we consider a prior knowledge base of only the road network for connecting county nodes, however, our framework provides flexibility towards incorporating other types of transportation networks.

For a given complementary intent pair of demander and supplier (d, s) for a resource category, where $d \in D$ and $s \in S$ and their hub nodes $d' \in I$ and $s' \in I$, we define the cost function as:

$$w_{ij} = C(d,s) = (1/K) * \sum_{x \in [1,K]} PathDistance(d', s', x) \quad (1)$$

where x is an alternative shortest path for the hub node pair (d', s').

For computing alternative paths for a given pair of nodes d and s, we employ Yen's k-shortest loopless paths algorithm [9] to find paths between their abstracted hub nodes d' and s' in the transport network. Taking into account only distance, however, can yield very similar shortest paths, since substituting some edges in the shortest path can generate multiple short paths. To avoid this and ensure that our paths are resilient, we use the following modification: each time we compute a shortest path between two counties, we eliminate all edges constituting that particular path from the network and re-run the Yen's algorithm for that pair of counties, in an iterative manner. This approach provides edge-disjoint, non-intersecting short paths. The $PathDistance(d', s', x)$ function for a path x is the total road distance in the county network for traveling over this path. Such a cost function can be computed for nodes in a national transport network of counties/cities offline, prior to a disaster event.

To be practical, for a disaster relief situation, we only consider paths between the source and destination hub nodes such that both belong to either the same state or adjacent states, i.e., within one hop of each other.

2.3 Path Connectivity Matrix Creation

We define a connectivity matrix for modeling uncertainty and risk in achieving connectivity between the hub nodes during a disaster event. Given that a disaster event can damage transportation paths, a supplier or demander would trade off reachability to minimize risk. We compute the likelihood of connecting any two adjacent county/hub nodes in the transport network by relying on the statistics of number of highways intersecting in each of the two county regions. Such an information can be learned from available offline knowledge bases of a nation's transport network, and can also be modified in an on-going disaster as the infrastructure damage information becomes available. Assuming that the amount of damage is independent of the transport network in each county node, we specifically define the connectivity likelihood for an adjacent pair of county nodes $i1$ and $i2$ as:

$$\Pr(i_1, i_2) = HighwayExistence(i_1) * HighwayExistence(i_2) \quad (2)$$

where for a county node i_a with H number of highways, the $HighwayExistence$ function computes the likelihood of existence of those highways during a disaster to enable the county's connectivity. It captures the risk to measure the availability of a highway. We define the likelihood of such availability of a highway H_g as $r \in [0, 1]$ (0.5 by default, but can be estimated by historical data of a road availability). We specifically define the probabilistic function for the overall availability of highways for the county as:

$$HighwayExistence(i_a) = 1 - (1 - r)^H \quad (3)$$

For the non-adjacent pair of county nodes i_a and i_b, we compute the joint probability of connectivity likelihood for the constituent pairs of adjacent county nodes, by considering the independence assumption for the connectivity risk between any two county pairs, given the possibility of independent effects on transport pathways by a disaster:

$$Pr(i_a, i_b) = (1/K) * \sum_{x \in [1,K]} \prod_{i_u, i_v \in \text{adjacent nodes on path x}} \Pr(i_u, i_v) \quad (4)$$

where x is an alternative shortest path for the hub node pair (d', s').

Table 1: Illustration of a local matching matrix for individual pairwise path selection, based on Figure 2

Alternative Path	Transportation Cost	Path Connectivity
Path-1	$PathDistance(d', s', s' - i_8 - i_1 - i_5 - d')$	$\Pr(s', i_8) * \Pr(i_8, i_1) * \Pr(i_1, i_5) * \Pr(i_5, d')$
Path-2	$PathDistance(d', s', s' - i_8 - i_3 - i_2 - i_6 - d')$	$\Pr(s', i_8) * \Pr(i_8, i_3) * \Pr(i_3, i_2) * \Pr(i_2, i_6) * \Pr(i_6, d')$
Path-3	$PathDistance(d', s', s' - i_8 - i_4 - i_6 - d')$	$\Pr(s', i_8) * \Pr(i_8, i_4) * \Pr(i_4, i_6) * \Pr(i_6, d')$

The proposed approach to defining *transportation cost* and *path connectivity* matrices allows for offline computation prior to a disaster, that is significant in contributing to near real time performance.

To our knowledge, this is the first approach in disaster informatics literature to combine such offline knowledge for online analytics on social media to assist disaster relief coordination. Furthermore, this approach is flexible to accommodate additional knowledge of a transport network edge metadata (e.g., speed limits), unavailability of an edge or disconnected network edges both during a disaster event as well as from the historical event data such as number of accidents, or number of times roads are unavailable due to floods.

2.4 Local Pairwise Path Selection and Global Matching Optimization

From the preceding subsections, we have two matrices to characterize the behavior of transport network before a disaster event. Using minimum transportation path cost and maximum path connectivity, we can now locally select the top q matches for an individual supplier node $s \in S$ against the set D of the bipartite intent network, based on the abstracted transport network of hub nodes as shown in Figure 2.

We compute path connectivity to characterize connectivity risk of an alternative path x between d' and s' (using Yen's algorithm modified for k diverse shortest paths as described in Section 2.2) in the transport network, by using functions in Equations 2 and 4. For example, in Figure 2, we could compute the local path matrix (*cf.* Table 1) for a local individual pair of nodes d and s (abstract nodes d' and s'), using alternative paths (e.g., *path-1*, *path-2*, and *path-3*). Such a matrix is computed for all possible pairs for s' and d'.

Given a set of such local individual path matrices for node pairs, we employ a *Pareto Optimal* strategy [2] to identify a ranked list of paths with minimum transportation cost but maximum likelihood for the path connectivity, within each local path matrix. We then rank across the local path matrices of a given supplier node s with all the possible nodes d related to a resource category c_l.

Following the Pareto Optimality criterion, we define a global matching recommendation list for the set of pairing edges between $s \in S$ and $d \in D$ (abstract nodes in S' and D') related to a resource

Table 2: Performance of our proposed matching approach in comparison with baselines for Hurricane Sandy

Algorithm	Avg path length between demand & supply locations	Avg path existence probability
Hungarian	491.1 km	0.79
APSP	465.73 km	0.823
Our method	518.585 km	0.925

category c. To select *topq* number of pairs with maximum connectivity and minimum transportation cost as well as one-one mapping for d and s, we define the following selection function:

$$CategoricalMatchSelection(D', S', c_l, Q) = \{\max_{\forall(s'',d'')|s''\in S'',d'\in D''} Pr(s'', d'')\} \tag{5}$$

where S'' and D'' sets are extracted from

$$\{\min_{\forall(s',d')|s'\in S',d'\in D'} PathDistance(d', s', x) \; \forall\, x \in \text{alternative paths of the pair (d',s')}\} \tag{6}$$

3 EXPERIMENTAL SETUP AND EVALUATION

For a robust validation of the proposed *Matching* framework for near realtime efficiency, we evaluated it on diverse disaster events. However, in the interests of space, we only present the evaluation results on 4.9 million tweets from a single event, *Hurricane Sandy* (October 27-November 7, 2012). We adapted an intent classifier for the disaster domain [5] to collect relevant tweets for our problem. It identifies the type of a tweet as a request for help (seeking intent), an offer (offering intent) to help or neither, which helped to construct our bipartite intent network. For creating the transport network, we used the data provided by US National Highway Planning Network with counties (hubs) as nodes, and edges based on the adjacency of counties in the physical space. We assign distances as well as probability of existence of paths during the disaster between every pair of demand and supply hub (county) nodes as outlined in Section 2. We use as baselines the following: (i) the $O(n^3)$ Edmonds-Karp version of the Hungarian algorithm [4], a combinatorial optimization algorithm used to solve general task assignment or transportation problems; and (ii) the $O(n^3)$ Floyd-Warshall algorithm that finds the shortest paths between two nodes (APSP).

Table 2 shows the performance of our proposed matching approach against two baselines. Both these baselines only minimize the path length, and do not take into account the probability of path existence. We observe that our method computes paths whose lengths are about 10% longer than the optimal path length (from APSP) between a given pair of locations. However, the paths suggested by our approach are at least 10% more reliable (in terms of existence probability) than the baselines.

We next study how the computation time of our approach varies with a variation in the number of demand and supply hub nodes, by sampling different numbers of hub locations for Hurricane Sandy. The likelihood of sampling hub locations is based on their overall frequency distribution (i.e., the number of supply offers or demand requests at that particular hub), which we inferred from the real-world Twitter data during the hurricane event. We sample x demand locations based on the demand frequency distribution and x supply locations based on the supply frequency distribution and compute the optimal matching between the demand and supply locations using our approach, and the baseline Hungarian algorithm. We observe from Figure 3 that our method can match new demand or supply requests that arise over time with minimum overheads. It is at least 10 times faster than the Hungarian algorithm and shows potential for usability in disaster relief situations.

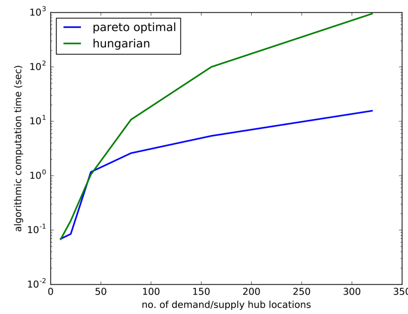

Figure 3: Execution time of our proposed pareto optimal approach in comparison with the Hungarian baseline

4 CONCLUSION

This paper presented a novel SMF model, for the online matching of users who seek help with those who offer help during a crisis situation. We frame this as a multi-objective optimization problem and propose an efficient, pareto-optimal strategy that models and optimizes the dual metrics of transportation path distances between help-seeking and help-offering locations, as well as the uncertainty of path existence during a disaster. In future, we will test the SMF model on more events and formalize additional uncertainties in it.

5 ACKNOWLEDGEMENT

This work is partially supported by the U.S. National Science Foundation grants EAR-1520870 and IIS-1657379.

REFERENCES

[1] M.M. Islam, J.V. Vate, J. Heggestuen, A. Nordenson, and K. Dolan. 2013. Transforming in-kind giving in disaster response: A case for on-line donation registry with retailers. In *IEEE GHTC*.
[2] H. Kung, F. Luccio, and F.P. Preparata. 1975. On finding the maxima of a set of vectors. *Journal of the ACM (JACM)* (1975).
[3] P. Meier. 2015. *Digital humanitarians: how big data is changing the face of humanitarian response*. Crc Press.
[4] J. Munkres. 1957. Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial & Applied Mathematics* (1957).
[5] H. Purohit, C. Castillo, F. Diaz, A. Sheth, and P. Meier. 2013. Emergency-relief coordination on social media: Automatically matching resource requests and offers. *First Monday* (2013).
[6] H. Purohit, M. Dalal, P. Singh, B. Nissima, et al. 2016. Empowering Crisis Response-Led Citizen Communities: Lessons Learned from JKFloodRelief. org Initiative. In *Strategic Management and Leadership for Systems Development in Virtual Spaces*.
[7] K. Starbird and L. Palen. 2011. Voluntweeters: Self-organizing by digital volunteers in times of crisis. In *ACM SIGCHI*.
[8] I. Varga, M. Sano, K. Torisawa, C. Hashimoto, K. Ohtake, T. Kawai, J. Oh, and S. De Saeger. 2013. Aid is Out There: Looking for Help from Tweets during a Large Scale Disaster.. In *ACL*.
[9] J.Y. Yen. 1971. Finding the k shortest loopless paths in a network. *Management Science* (1971).