**MURI yearly progress report template**

Reporting period: August 31, 2015 -- August 31, 2016

Name: Zhi-Li Zhang

Papers:

DOI (if known): 10.1007/978-3-319-42345-6\_2

Report type: conference paper

Authors: Braulio Dumba, Golshan Golnari, and Zhi-Li Zhang

Title: *Analysis of a Reciprocal Network using*

*Google+: Structural Properties and Evolution*

Keywords: social networks, reciprocity, reciprocal networks, evolution, Google+

Abstract

 Many online social networks such as Twitter, Google+, Flickr

and Youtube are directed in nature, and have been shown to exhibit a

nontrivial amount of reciprocity. Reciprocity is de\_ned as the ratio of the

number of reciprocal edges to the total number of edges in the network,

and has been well studied in the literature. However, little attention is

given to understand the connectivity or network form by the reciprocal

edges themselves (reciprocal network), its structural properties, and how

it evolves over time. In this paper, we bridge this gap by presenting a

comprehensive measurement-based characterization of the connectivity

among reciprocal edges in Google+ and their evolution over time, with

the goal to gain insights into the structural properties of the reciprocal

network. Our analysis shows that the reciprocal network of Google+ re-

veals some important user behavior patterns, which reect how the social

network was being adopted over time

DOI (if known):

Report type: conference paper (submitted)

Authors: Braulio Dumba and Zhi-Li Zhang

Title: Unfolding the Core Structure of the Reciprocal

Graph of a Massive Online Social Network

Keywords: social networks, Reciprocity, Reciprocal Network, Google+, Network Core Structure

Abstract

Google+ (G+ in short) is a directed online social network

where nodes have either reciprocal (bidirectional) edges or parasocial

(one-way) edges. As reciprocal edges represent strong social ties, we study

the core structure of the subgraph formed by them, referred to as the

reciprocal network of G+. We develop an effective three-step procedure

to hierarchically extract and unfold the core structure of this reciprocal

network. This procedure builds up and generalizes ideas from the existing

k-shell decomposition and clique percolation approaches, and produces

higher-level representations of the core structure of the G+ reciprocal

network. Our analysis shows that there are seven subgraphs (“communities") comprising of dense clusters of cliques lying at the center of the

core structure of the G+ reciprocal network, through which other com-

munities of cliques are richly connected. Together they form the core to

which “peripheral" sparse subgraphs are attached.

DOI (if known): [10.1145/2935755.2935760](http://dx.doi.org/10.1145/2935755.2935760)

Report type: conference paper

Authors: Arvind Narayanan, Saurabh Verma, Zhi-Li Zhang

Title: Mining Spatial-Temporal geoMobile Data via Feature Distributional Similarity Graph

Keywords: spatial-temporal data analysis, geoMobile datasets, feature similarity graph, mobility, spectral clsutering

Abstract

Mobile devices and networks produce abundant data that exhibit geo-spatial and temporal properties mainly driven by human behavior and activities. We refer to such data as geoMobile data. Mining such data to extract meaningful patterns that are reflective of collective user activities and behavior can benefit mobile network resource management

as well as the design and operations of mobile applications

and services. However, diverse feature distributions inherent

in such data make such a task challenging. In this paper we

advocate an approach based on advanced machine learning

algorithms to transform original data matrices into a feature

distributional similarity graph and extract “latent" patterns

from complex structures of geoMobile data. Our analysis is

further aided with a visualization technique. Using mobile

access data records from an operational cellular carrier, we

demonstrate the potentials of our proposed approach under

multiple settings, and make some very interesting observations from the obtained results.

DOI (if known):

Report type: conference paper

Authors: Arvind Narayanan, Saurabh Verma, Zhi-Li Zhang

Title: Most Calls are Local (but Some are Regional): Dissecting Cellular Communication Patterns

Keywords: data analytics, cellular communication patterns, Laplaicain Eigenmaps, visualization

Abstract

 We conduct a detailed analysis of cellular communication patterns using (voice/text based) call detail records (CDR) dataset from a nationwide cellular network. We analyze a 5-month large dataset containing over hundreds of millions of CDRs with a user population of over 5 million to dissect meaningful communication patterns, with the goal to understand their impact on – and better manage – cellular network resources. What makes this dataset interesting is that we have both location and timestamp information of the caller and the callee. This allows us to associate communication patterns of users with geographic locations. The enormous size and diversity inherent in the (big) data set, however, makes extracting communication patterns a challenging task. We illustrate this diversity by analyzing tower-level activities and communication patterns between towers and find certain patterns emerging. However, due to the complex structure of the data, extracting them becomes non-trivial. By providing structures to the data in the form of matrices, we adopt machine learning techniques (Laplacian Eigenmaps and t-SNE) to extract “latent” patterns from the data, while accounting for the inherent non-linearity and skewed data distributions. Our main results reveal the existence of interesting regional communication patterns of varying localities and sizes, out of which one pattern scatters across the entire nation. Last but not the least, we also find a number of distinct communication patterns co-existing within the capital city of the nation.

DOI (if known): [10.1109/INFOCOM.2016.7524508](http://dx.doi.org/10.1109/INFOCOM.2016.7524508)

Report type: conference paper

Authors: Cheng Jin, Abhinav Srivastava, Zhi-Li Zhang

Title: Understanding Security Group Usage in a Public IaaS Cloud

Keywords: mobile cloud, Infrastructure-as-a-Service (IaaS), Openstack, security group

Abstract

To ensure security, cloud service providers employ security groups as a key tool for cloud tenants to protect their virtual machines (VMs) from attacks. However, security groups

can be complex and often hard to configure, which may result

in security vulnerabilities that impact the entire cloud platform.

The goal of this paper is to investigate and understand how cloud

tenants configure security groups and to assist them in designing

better security groups. We first conduct a measurement-based

analysis of security group configuration and usage by tenants in

a public IaaS cloud. We then propose and develop a tool called

Socrates, which enables tenants to visualize and hence understand

the static and dynamic access relations among VMs. Socrates

also helps diagnose potential misconfigurations and provides

suggestions to refine security group configurations based on

observed traffic traversing tenants’ VMs. Applying Socrates to all

tenants hosted on the public IaaS cloud, we analyze the common

usage (“good” as well as “bad” practices) of cloud security groups

and report the key lessons learned in our study. To the best of our

knowledge, our work is the first to analyze cloud security group

usage based on real-world datasets, and to develop a system to

help cloud tenants understand, diagnose and better refine their

security group configurations.

DOI (if known): [10.1109/INFOCOM.2016.7524600](http://dx.doi.org/10.1109/INFOCOM.2016.7524600)

Report type: conference paper

Authors: Yang Zhang, Hesham Mekky, Zhi-Li Zhang, Fang Hao, Sarit Mukherjeeand T V Lakshman

Title: SAMPO: Online Subflow Association for Multipath TCP with Partial Flow Records

Keywords: multiple-path TCP (MPTCP), MPTCP subflow detection, network measurement, sampling

Abstract

Multipath TCP (MPTCP) is a promising technique

for boosting application throughput while using well-known and

versatile network socket interfaces. Recently, many interesting

applications of MPTCP in various environments such as wireless

networks and data centers have been proposed, but little work

has been done to investigate the impact of this protocol on

conventional network devices. For example, MPTCP throughput

advantage can be better achieved if all MPTCP subflows are

routed on disjoint paths, but this is currently not feasible since

routers are not designed to recognize the membership of MPTCP

subflows. In this paper, we take a first step to address this issue by

proposing SAMPO, an online algorithm to detect and associate

MPTCP subflows in network. The main challenge is that sampling

techniques and network dynamics may cause a network device to

only obtain partial flow records. SAMPO takes advantage of both

protocol information and statistical characteristics of MPTCP

data sequence number to overcome the challenge in network.

Through analysis and experimentation, we show that SAMPO is

able to detect and associate MPTCP subflows with high accuracy

even when a small portion of the entire flow records are available

DOI (if known): 10.1007/s10922-016-9373-0

Report type: journal paper

Authors: Braulio Dumba, Hesham Mekky, Sourabh Jain, Guobao Sun & Zhi-Li

Zhang

Title: A Virtual Id Routing Protocol for

Future Dynamics Networks and Its Implementation Using the SDN Paradigm

Keywords: VIRO, Open vSwitch, Software Defined Networks, GENI

Abstract

In this paper, we propose Virtual Id Routing (VIRO) a novel ‘‘plug-&-

play’’ non-IP routing protocol for future dynamics networks. VIRO decouples

routing/forwarding from addressing by introducing a topology-aware, structured

virtual id layer to encode the locations of switches and devices in the physical

topology. It completely eliminates network-wide flooding in both the data and

control planes, and thus is highly scalable and robust. VIRO effectively localizes the

effect of failures, performs fast re-routing and supports multiple (logical) topologies

on top of the same physical network substrate to further enhance network robustness.

We have implemented an initial prototype of VIRO using Open vSwitch, and

we extend it (both within the user space and the kernel space) to implement VIRO

switching functions in VIRO switches. In addition, we use the POX SDN controller

to implement VIRO’s control and management plane functions. We evaluate our

prototype implementation through emulation and in the GENI (the Global Environment

for Network Innovations) testbed using many synthetic and real topologies.

Our evaluation results show that VIRO has better scalability than link-state based protocols (e.g. OSPF and SEATTLE) in terms of routing-table size and control

overhead, as well as better mechanisms for failure recovery.

DOI (if known): 10.1002/nem.1900

Report type: journal paper

Authors: Yang Zhang, Hesham Mekky, Zhi-Li Zhang,\_ Ruben Torres, Sung-Ju Lee, Alok Tongaonkar and Marco Mellia

Title: Detecting malicious activities with user-agent-based profiles

Keywords: Network Security; HTTP User-Agent; Flow Correlation, Malicious Activity Detection

Abstract

Hypertext transfer protocol (HTTP) has become the main protocol to carry out malicious activities. Attackers typically use HTTP for communication with command-and-control servers, click fraud, phishing and other malicious activities, as they can easily hide among the large amount of benign HTTP traffic. The useragent (UA) field in the HTTP header carries information on the application, operating system (OS), device, and so on, and adversaries fake UA strings as a way to evade detection. Motivated by this, we propose a novel grammar-guided UA string classification method in HTTP flows. We leverage the fact that a number of ‘standard’ applications, such as web browsers and iOS mobile apps, have well-defined syntaxes that can

be specified using context-free grammars, and we extract OS, device and other relevant information from them. We develop association heuristics to classify UA strings that are generated by ‘non-standard’ applications that do not contain OS or device information. We provide a proof-of-concept system that demonstrates how our approach can be used to identify malicious applications that generate fake UA strings to engage in fraudulent activities.

Papers presented at meetings but not published

Numbers

Title, authors, details

(Send copies)

Papers presented at meetings, but not published in conference proceedings

None

Honors and Awards

Number of patents submitted

List of Patent Titles Submitted

Number of Patents Awarded

List of Patent Titles awarded

Scientific progress and accomplishments (Description should include significant theoretical or experimental advances) Plain Text Only - If you have diagrams, formula's, etc. in a Word, PDF or other document, enter "See Attachment" below and use the "Attachment" section at the bottom of the menu. This section can include, but is not limited to: (1) Foreword (optional) (2) Table of Contents (if report is more than 10 pages) (3) List of Appendixes, Illustrations and Tables (if applicable) (4) Statement of the problem studied (5) Summary of the most important results (6) Bibliography (7) Appendixes

Send me your individual summaries and I’ll assemble something coherent for the ARO site.

(1) Forward

In the fourth year of this collaborative MURI project, one major research direction that we have carried is on analysis and evolution of the *reciprocal* network – a graph formed by nodes with reciprocal edges (i.e., mutually connected nodes with bi-directional connections) arising from a directed online social network. For this study, we in particular leveraged the datasets we have collected from the social network Google+, we studied the structural properties of the reciprocal network of Google+ and its evolution. As part of our ongoing work, we are study the core structure of the reciprocal network and its formation. We have also continued to analyze the potential emergences of “*multivariate heavy-tail*” phenomena in various complex networks and their roles in shaping the *structural* properties of these networks.

Another major direction we have focused on analyzing and extracting actionable information from geoMobile data. We coin the term “geoMobile” data to refer to spatial-temporal datasets arising from complex networks which exhibit geo-spatial and temporal properties mainly driven by human behavior and activities. We have developed a novel framework based on advanced machine learning techniques to transform original data matrices into a feature distributional similarity graph and extract “latent" patterns from complex structures of geoMobile data. As an application of this framework, we have conducted a detailed analysis of cellular communication patterns using (voice/text based) call detail records (CDR) dataset from a nationwide cellular network. We analyzed a 5-month large dataset containing over hundreds of millions of CDRs with a user population of over 5 million to dissect meaningful communication patterns, with the goal to understand their impact on – and better manage – cellular network resources.

Furthermore, as applications of multivariate heavy tail analysis, we have conducted several studies on cyber (cloud, network and web) security, including understanding security group usage in a public mobile IaaS Cloud, online subflow association for multipath TCP with partial flow records as well as detecting malicious activities with user-agent-based profiles in HTTP web traffic.

In the following, we briefly summarize the major scientific progress and accomplishments we have made in the fourth year of this project.

(4) Statement of the Problem

**Analysis of a Reciprocal Network using Google+: Structural Properties and Evolution:** Many online social networks are fundamentally directed : they consist of both

reciprocal edges, i.e., edges that have already been linked back, and parasocial

 edges, i.e., edges have not been or is not linked back [1]. Reciprocity is de\_ned

as the ratio of the number of reciprocal edges to the total number of edges in the

network. It has been shown that major online social networks (OSN) that are

directed in nature, such as Twitter, Google+, Flickr and Youtube, all exhibit a

nontrivial amount of reciprocity: for example, the global reciprocity of Flickr,

Youtube, Twitter and Google+ have been empirically measured to be

0.62, 0.79, 0.22 and 0.32, respectively. Reciprocity has been widely studied in

the literature.

Reciprocal edges represent the most stable type of connections or relations in

directed network -- they reflect strong ties between nodes or users, such

as (mutual) friendships in an online social network or \following" each other in a

social media network like Twitter. Connectivity among reciprocal edges can thus

potentially reveal more information about users in such networks. For example,

a clique formed by reciprocal edges suggest users involved are mutual friends or

share common interests. More generally, it is believed that nontrivial patterns in

the reciprocal network -- the bidirectional subgraph of a directed graph could reveal possible mechanism of social, biological or different nature that systematically acts as organizing principles shaping the observed network topology. Moreover, understanding the dynamic structural properties of the reciprocal network can provide us with additional information to characterize or compare directed networks that go beyond the classic reciprocity metric, a single static value currently used in many studies. However, little attention has been paid in the literature to understand the connectivity between reciprocal edges -- the reciprocal network -- and how it evolves over time. In this work, we have performed a comprehensive measurement-based characterization of the connectivity and evolution of reciprocal edges in Google+, in order to shed some light on the structural properties of G+'s reciprocal network.

**Mining Spatial-Temporal geoMobile Data via Feature Distributional Similarity Graph:** Mobile technology has revolutionized the way how we live and interact with each other and the physical world. These mobile devices -- together with the applications running on top and the underlying mobile infrastructures supporting them -- have also enabled us to collect a whole gamut of (spatial-temporal) data that were possible otherwise. Combining such data arising from the cyber and/or physical worlds

with social data arising from human actions and behavior allows us to gain a deeper understanding of the events and changes occurring in our environs. The real challenge to-

day however is in designing methods and applying advanced analytics to extract actionable knowledge from such data and enable users to make sound and smart decisions. We coin the term *geoMobile* data to refer to datasets characterized by two salient features: i) they are associated with geo-locations (e.g. gathered by cell towers), and, ii) more often they capture human actions while on-the-move. For example, GPS-enabled smartphones produce feature-rich geoMobile datasets capturing human mobility and interactions with high spatial and/or temporal resolution. Evaluating human mobility can help unravel how populations move across regions. Human populations and their movements help understand carbon footprints and also aid in building smart grid energy networks by understanding the dynamically changing energy demands. With abundance of such geoMobile data, extracting meaningful patterns is an important yet non-trivial data analysis task that has wide applications, from network traffic engineering to urban transportation management, smart city planning, social behavior analysis, developing personalized services, geo-targeted mobile ads and cyber-physical world security.

In this work, we aim to unravel underlying latent structures from a (spatial-temporal) geoMobile dataset driven by human mobility and behavior. Our dataset contains user-generated data (or Internet) access records registered by an operational cellular service provider covering approximately a 7000 sq. mile region in the San Francisco bay area. We try to ask questions such as: what is the impact of human mobility over cellular base-station (or tower) activities? If such an impact exists, is it diverse? If yes, is there a way to identify them from a large cellular network with millions of users and close to a thousand towers? Along with mobility, does time-of-the-day (temporal aspects) affect tower-level activities? If yes, is the nature of impact same for all towers? If not, how

do we identify towers having similar impact due to human mobility and temporal factors? We believe understanding such underlying patterns driven by human mobility are vital to obtain insights about the events and changes that take place in the surrounding environment. For example, under future 5G deployments, cellular service providers can get insights about tower-level activities to wheel out strategies for load balancing, dynamic provisioning of resources and transport layer networks, and other related services.

**Most Calls are Local (but Some are Regional): Dissecting Cellular Communication Patterns:** Wide adoption of smart phones and other mobile devices has led to rapid growth in mobile traffic, which places a huge demand on the cellular network infrastructure such as resources on cell towers, radio network controllers, and so forth. Gaining a deeper understanding of cellular usage patterns and how they are affected by user behavior and mobility is critical to effective management of cellular network resources and to meet user quality of experience expectation. Originally designed for billing purpose, the

Call Detail Records (CDRs) collected by cellular network

operators provide a useful and rich data source for obtaining

insights into network usage patterns and user behavior. Since

CDRs are collected at either an initiating cell tower or a

terminating cell tower or both (when both caller and callee

belong to the same cellular service provider), they allow for

more detailed studies of cellular network usage patterns at

the (finer-grained) cell tower level2 . In addition, since CDRs

are typically stored by cellular network operators for longer

periods of time (e.g., for billing), one can also conduct

studies of cellular usage patterns over a longer time horizon

resulting in the generation of big datasets.

In this work, we have studied the cellular communication patterns at the cell tower level using the CDRs collected by a national cellular provider over a period of close to five months. In particular, using voice calls and SMS messages (thereafter we refer to both simply as “calls”) – originating and terminating at cell towers within the same nation-wide cellular provider, we construct the nation-wide (and time varying) traffic matrices at the cell tower level. From a network-wide perspective, we leverage the (tower-level) traffic matrices to analyze the usage patterns and geographical distributions

of calls at individual cell towers as well as across origin-destination cell tower pairs. We find that call volumes at cell towers are highly diverse, and vary drastically from towers to towers with strong geographical effects. In many towers, a large portion of calls originate and terminate at the same towers, indicating both callers and callees reside within a local area covered by a single tower. On the other hand, the geographical coverage and density distributions of cell towers are highly skewed, with many more towers in large urban areas. While there are strong correlations between geographical distances and call volumes at cell tower level, locality can only explain part of the communication patterns apparent in the cellular call data.

(5) Summary of the Most Important Results:

**Analysis of a Reciprocal Network using Google+: Structural Properties and Evolution:** In this work, we have performed a comprehensive measurement-based characterization of the connectivity and evolution of reciprocal edges in Google+, in order to shed some light on the structural properties of G+'s reciprocal network. We are particularly interested in understanding how the reciprocal network of G+ evolves over time as new users (nodes) join the social network, and how reciprocal edges are created, e.g., whether they are formed mostly among extant nodes already in the system or by new nodes joining the network. We summarize the major findings of our study as follows:

* We find that the density of Google+ -- which reflects the overall degree of social connections among Google+ users -- decreases as the network evolves from its second to third year of existence. This finding differs from the observations, where it was reported that that Google+ social density fluctuates in an increase-decrease fashion in three phases, but it reaches a steady increase in the last phase during its first year of existence.
* Furthermore, we observe that both the density and reciprocity metrics of Google+'s reciprocal network also decrease over time. Our analysis reveals that these are due to the fact that the new users joining Google+ later tend to be less “social" as they make fewer connections in general. In particular, i) the number of users creating at least one reciprocal edge is decreasing as the network evolves; ii) the new users joining the reciprocal network are creating fewer edges than the users in the previous generation.
* We show that if a user does not create a reciprocal edge when he/she joins Google+, there is a lower chance that he/she will create one later. In addition, users who already have reciprocal connections with some users tend to create more reciprocal connections with additional users.

To the best of our knowledge, our study is the first study on the properties and evolution of a “reciprocal network" extracted from a *directed* social graph.

**Mining Spatial-Temporal geoMobile Data via Feature Distributional Similarity Graph:** Using anonymized data access records from an operational cellular network as a case study, we argued that classical PCA and related linear methods are ineffective in mining diverse geoMobile data, due to inherent high variability, non-linearity and skewed feature distributions. To address these challenges, we have proposed a novel approach where we first transform original data matrices into a *feature distributional similarity graph*, and then apply spectral clustering to project the feature space into lower

dimensional manifolds for latent significant feature extraction. To help visualize the resulting clusters, we leverage a density preserving mapping tool to embed the “now" labeled data points in a two (or three) dimensional space. We demonstrated the potential of this approach by applying it to our dataset under multiple settings and observed some insightful patterns. As ongoing research, we are taking a deeper look to understand the underlying theoretical guarantees of our proposed approach, and develop mechanisms to characterize the extracted clusters in order to interpret the results better.

**Most Calls are Local (but Some are Regional): Dissecting Cellular Communication Patterns:** In this work, we presented our detailed analysis on a nationwidelarge CDR dataset enriched with both source anddestination cellular tower information of over 500 millioncalls and found some interesting patterns. To dissectand extract the significant latent communication patterns inthe call traffic matrices, we apply the Laplacian Eigenmapmethod by generating a (high-dimensional) similarity matrix from an origin-destination (OD) call matrix based on theempirical distributions of calls from one tower to other towers. This allows us to account for the highly diverse data distributions in the original call traffic matrix, and allows us to extract latent patterns or “clusters” lying in certain lower-dimensional (non-linear) manifolds. We also developa visualization tool to illustrate and interpret the extracted communication patterns. The main findings of our study are summarized below:

* Although 25% of the total call volumes are generated and consumed by the towers in and around the capital of the nation, the nature of calls, such as incoming versus outgoing volumes vary markedly across towers and regions. Not surprisingly, there are strong dependencies between call volumes at tower levels and human activities around these towers.
* A general observation is that for most of the towers, a majority of the calls are local: namely, the caller and the callee for most calls are associated with the same tower, or a conglomerate of cellular towers that are located geographically within close proximity. However, we show that such locality effects are diverse in that the geographical boundaries are not clear-cut and cannot be simply defined based on geographic distance alone.
* Since our empirical analysis suggests the existence of certain patterns emerging out of the communication between towers, we briefly describe an approach to unravel such hidden patterns by using recently developed state-of-the-art machine learning tools. We find that most of the communication patterns are regional with varying localities and sizes. Moreover, we find that even within the capital city, there are a number distinct “regional” communication patterns, suggesting the presence of vast diversity in social interactions and human behavior. There is also one communication pattern containing towers that are sparsely distributed across the nation, many of which are located along major transportation networks; this pattern likely captures call activities of long-distance travelers.

 (6) Bibliography

## Braulio Dumba, Golshan Golnari, and Zhi-Li Zhang, *Analysis of a Reciprocal Network using Google+: Structural Properties and Evolution,* In Proc. of the 5th International Conference Computational Social Networks Conference (CSoNet 2016), Ho Chi Minh City, Vietnam, August 2-4, 2016, Volume 9795, Lecture Notes in Computer Science, pp 14-26.

## Arvind Narayanan, Saurabh Verma, Zhi-Li Zhang, *Mining Spatial-Temporal geoMobile Data via Feature Distributional Similarity Graph,* In *Proceedings of the First ACM Workshop on Mobile Data* (MobiData'16), 13-18, June 30, 2016, Singapore, DOI: <http://dx.doi.org/10.1145/2935755.2935760>

## Arvind Narayanan, Saurabh Verma, Zhi-Li Zhang, *Mining Spatial-Temporal geoMobile Data via Feature Distributional Similarity Graph,* To appear in *Proceedings of IEEE Globecom 2016,* December 4-8, 2016, Washington DC.

## Cheng Jin, Abhinav Srivastava, Zhi-Li Zhang, *Understanding Security Group Usage in a Public IaaS Cloud*, In Proc. of The 35th Annual IEEE International Conference on Computer Communications (IEEE INFOCOM 2016), San Francisco, CA, USA, 2016. doi: 10.1109/INFOCOM.2016.7524508

1. Yang Zhang, Hesham Mekky, Zhi-Li Zhang, Fang Hao, Sarit Mukherjeeand T V Lakshman, *SAMPO: Online Subflow Association for Multipath TCP with Partial Flow Records*, In Proc. of The 35th Annual IEEE International Conference on Computer Communications (IEEE INFOCOM 2016), San Francisco, CA, USA, 2016. doi: 10.1109/INFOCOM.2016.7524600
2. Braulio Dumba, Hesham Mekky, Sourabh Jain, Guobao Sun and Zhi-Li Zhang, *A Virtual Id Routing Protocol for Future Dynamics Networks and Its Implementation Using the SDN Paradigm*, Journal of Network System Management (2016) 24: 578. doi:10.1007/s10922-016-9373-0
3. Yang Zhang, Hesham Mekky, Zhi-Li Zhang,\_ Ruben Torres, Sung-Ju Lee, Alok Tongaonkar and Marco Mellia, *Detecting malicious activities with user-agent-based profiles*, International Journal of Network Management,, 25: 306–319. doi: [10.1002/nem.1900](http://dx.doi.org/10.1002/nem.1900).
4. Braulio Dumba and Zhi-Li Zhang**,** *Unfolding the Core Structure of the Reciprocal Graph of a Massive Online Social Network****,***  Submitted to COCOA’16

**Technology Transfer** (any specific interactions or developments which would constitute technology transfer of the research results). Examples include interaction with other DOD scientists, interactions with industry, initiation of a start-up company based on research results or transfer of information which might impact the development of products. Plain Text Only - If you have diagrams, formula's, etc. in a Word, PDF or other document, enter "See Attachment" below and use the "Attachment" section at the end of the menu for your document.

**Faculty**

Name, % supported, National Academy Member?

(% supported is percent (0-1) of total time person was supported by this grant for the reporting period)

**Zhi-Li Zhang, 2 months**

grad students

Name, % supported, discipline of degree

**Golshan Golnari, 9 months, ECE**

**Braulio Dumba, 9 months, CS**

**Saurabh Verma, 6 months, CS**

Post docs

Name, % supported, discipline of degree

Master Degrees Awarded

Name (% supported is percent (0-1) of total time person was supported by this grant)

Undergraduates

Name, % Supported, discipline of degree

**Peter Kneller, 1 month, mathematics**

Student Metrics

|  |
| --- |
| **All items refer to graduating undergraduates funded by this agreement and the reporting period for this report.** |

|  |  |
| --- | --- |
| Number of graduating undergraduate students: | 0 |
| Number of undergraduate students graduating with degrees in science, mathematics, engineering, and technology fields:  | 0 |
| Number of graduating undergraduates who will continue to pursue graduate degrees: | 0 |
| Number of graduating undergraduates who intend to work for the Defense Department: | 0 |
| Number of graduating undergraduates during this period who achieve a 3.5 to 4.0 GPA (Convert GPAs on any other scale to be an equivalent value on a 4.0 scale.): | 0 |
| Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering: | 0 |
| The number of undergraduates funded by your agreement who graduatedduring this period and will receive scholarships or fellowships forfurther studies in science, mathematics, engineering or technology fields.: | 0 |

Doctorate Degrees Awarded

Name (% supported during reporting period)

Other staff

Name (% supported)

Any change in your forcast expenditures?