Emmanouil Papangelis (Manos Papagelis)

2919 1/2B Shattuck Ave, Berkeley, CA, 94705, United States
papaggel@berkeley.edu, *+1.647.995.8500* (cell)
http://www.ischool.berkeley.edu/people/faculty/manospapagelis

I am currently a *postdoctoral research scholar* at the University of California, Berkeley's School of Information (iSchool). I also serve in the *professional faculty* of UC Berkeley iSchool's professional Master of Information and Data Science (MIDS), delivered online. I hold a *PhD* [26] in Computer Science from the University of Toronto, Canada. From 2010 to 2015, I worked as a *Sessional Lecturer* at the Department of Computer Science, University of Toronto, Canada. Before going to Toronto, I received a *M.Sc.* [25] and a *B.Sc.* [24] in Computer Science from the University of Crete, Greece and worked as a *research fellow* at the Institute of Computer Science, FORTH, Greece. I spent the summers of 2009 and 2010 at Yahoo! Labs, working as a *research intern* with the *Social Media* and the *Usage Mining and Link Analysis* groups respectively.

My PhD research focused on the algorithmic and computational issues involved in the study of online social interactions and processes and contributed to (a) a better understanding of principles and models that govern social interaction and processes in online social media, and (b) to the design and development of algorithms, methods and tools that improve social interactions and processes online. The nature of these problems required an interdisciplinary research approach with questions and ideas flowing from both computing and social sciences. My general research interests span across the following research areas:

- Graph and Information Network Mining
- Data Mining and Analysis
- Databases and Knowledge Discovery
- City Science and Urban Informatics

My publication record reflects these interests. Most of this research work has been conducted jointly with my PhD advisor Prof. Nick Koudas, my M.Sc. advisor Prof. Dimitris Plexousakis, Dr. Martin Doerr at ICS-FORTH, Greece and my collaborators while at Yahoo! Research, Aris Gionis, Francesco Bonchi, Vanessa Murdock and Roelof van Zwol. The following sections provide an overview of my research work.

Google Scholar Citations Analysis

I am including below a summary of my citation analysis as compiled by the Google Scholar Citations service in Oct 2015. For more details and an updated chart please visit my public profile at http://goo.gl/3bi9m.

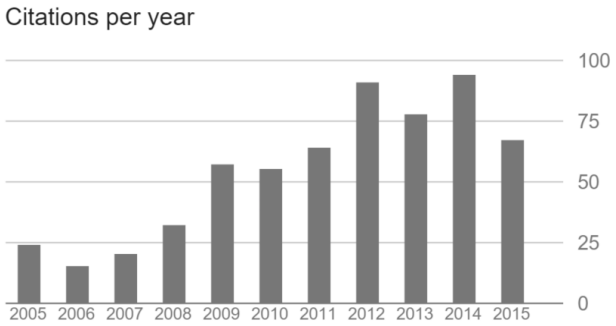

Figure 1: Google Scholar Citations Analysis: *citations*: **600+**, *h-index*: **11** (Oct'15)

1 Graph and Information Network Mining

1.1 Graph Augmentation

During my second research internship at Yahoo! Research, Barcelona, I became interested to the problem of suggesting a number of non-existing edges to add in a graph in order to optimize a connectivity problem. It is known that slight modifications in the network topology of a graph, might have a dramatic effect on its connectivity and thus to its capacity to carry on social processes, such as information cascades. We approached network modification as a graph augmentation problem where we seek to find a few non-existing edges to add to the graph. To this end, we presented methods that can accurately and efficiently evaluate the importance of non-existing edges to guide the graph augmentation process. In the augmented graph (i) social processes evolve faster and can reach more nodes, and (ii) random walks can converge a multitude times faster, giving rise to faster graph simulations. Preliminary results of this work have been published at the ACM CIKM 2011 conference [6]. More recent research results have appeard in the Hellenic Data Management Symposium (HDMS) [4] and have been published in the ACM Transactions on Knowledge and Discovery from Data journal (ACM TKDD) [15].

1.2 Social Influence and Diffusion in Online Social Media

In the presence of social influence, an idea, behavior norm, or product diffuses through the social network like an epidemic. Analysis of massive data sets at an aggregate level raises interesting observations of information cascades within social systems, but it has its limitations as it typically does not allow for interpretations at the individual level. More importantly, it does not allow to argue about the causality of the observed cascades. During my first internship at Yahoo! Labs, my research was driven by more refined research questions related to information cascades. Why cascades happen? How to determine whether cascades will occur in a social system? How one can influence a cascade process? Social scientists agree that in the presence of social influence, an idea or behavior diffuses through a social network like an epidemic. In our research, we were interested to develop methods that given a history of individual actions and social relationships of individuals can qualitatively and quantitatively detect social influence effects in a social system. Further, we developed a basic framework that allows to study the causality between individual actions of users and the social influence that they exert to their social network. As a case study for our research we employed a popular online social graph, Flickr (including 525K nodes, 47M edges), and our analysis of social influence was based on the cascade of the geotagging behavior. To the best of our knowledge, our work was one of the first that tried to bridge the gap between aggregate level cascade processes and individual behavior using large data sets. Results of this work have been published at the ACM Hypertext 2011 conference [10], while two U.S. patent applications have been issued based on this work [22, 23].

1.3 Analysis of Patterns of Communication in Social Media

With an increasing number of people that read, write and comment on blogs, the blogosphere has established itself as a dynamic medium of communication. In our research we are investigating patterns of communication in social media. In particular, we have conducted a thorough analysis of how information cascades in the blogosphere (i.e., the succession of linking behavior through which information propagates among blogs). Our analysis utilized one of the largest data sets available at the time (including 30M blogs and over 400M posts) and was distilled into a comprehensive report that contributed to a better understanding of the overall linking activity in the blogosphere. In particular, we analyzed: (i) trends on the degree of engagement and reaction of bloggers in stories that become available in blogs, (ii) the structure of cascades that are attributed to different population groups constrained by factors of gender, age, and continent, and (iii) topic-sensitive information cascades. Our analysis was the first to incorporate heterogeneity (differences in the groups) and revealed notable variations in the structural properties of cascades and in the ability of blogs to trigger prompt and widely-spread cascades that mainly depend on the blogger's profile and the topic of a post. Analyzing

information cascades can be useful in various domains, such as providing insight of public opinion and developing better prediction models with applications in health, business, politics and more. More importantly, our study established evidence of a cascading effect of online communication occurring in online social media and stimulated the next research questions. Results of this work were published at the 3rd Int. AAAI Conf. on Weblogs and Social Media (ICWSM 2009) [5].

2 Data Mining and Analysis

2.1 Social Search

The main idea of social search is to utilize social cues available in online social networks to improve search results, towards personalized search. An interesting, but challenging data mining task in (static and dynamic) social networks is given a specific user (or node) to probe the nodes (and/or the information they are associated with) that are found a few hops away from that user. We presented sampling-based algorithms that given a user in a social network can quickly obtain a near-uniform random sample of users in its neighborhood. Further, we employed these algorithms to quickly approximate the number of users in a user's neighborhood that have endorsed an item. Our methods are highly accurate and very efficient and can be utilized in a number of applications where we aim to rank items in a network, such as, to improve the quality of the user's search experience in a social search engine, or to improve the accuracy of collaborative filtering methods in a recommendation engine. Preliminary results of this work were published at the CIKM/SSM (Search in Social Media Workshop) [27]. More recent results have been published in the IEEE Transactions on Knowledge and Data Engineering journal [16].

2.2 Online Social Navigation

The primitive way to access all online information remains the web browser. But, web browsing or navigation is commonly assumed to be an autonomous and passive process where a user interacts with information but not with other people. Motivated by this fact we focused on ways to make web browsing more social. In our research, we considered a "social navigation" paradigm for browsing the web. Social navigation is the process where a number of people that share interests and searching goals decide to coordinate their efforts. With that in mind, we designed and developed a practical social navigation system that makes users aware of other users accessing similar information at similar time and encourages interpersonal real-time communication and collaboration. Preliminary results of this work were published at the ACM Hypertext Conference [3]. More recent results were published at the IEEE/WIC/ACM Web Intelligence Conference [2].

2.3 Recommendation Algorithms

Study of recommendation algorithms has been a long-term agenda item that has led to a number of publications and to the development of a recommendation system, which served both as a research platform and as a free online service. First we investigated the way in which recommendation algorithms can be employed in order to discover dynamic, virtual, online communities [11]. We spent the next period developing and evaluating the quality of collaborative filtering recommendation algorithms that are based on item similarities, instead of user similarities. The results appeared in the CIA Workshop [28] and after invitation to a special issue of the Elsevier Engineering Applications of Artificial Intelligence Journal [17]. Next, we turned focus on the scalability problem of recommendation algorithms. We developed a method of incremental computation of user similarities that was improving the performance of recommendation algorithms without reducing their quality. Results were published in the proceedings of the International Symposium on Methodologies for Intelligent Systems (ISMIS) [14]. Then, we investigated trust implications in web based social networks. We were particularly interested in modeling trust, developing a computational model for trust, and investigating the way in which trust fits architectonically into large scale information discovery systems that function within

highly-distributed environments. We argued that trust propagation techniques could be efficiently employed so as to alleviate the sparsity problem of recommendation systems. Results of this work were published at the International Conference on Trust Management (iTrust) [12].

3 Databases and Knowledge Discovery

3.1 Information Integration

While at ICS-FORTH, I was engaged in the "Ubi-Erat-Lupa" EU funded research project. The objectives of the project were to interlink archaeological research on the Roman era systematically and transnationally and to exchange expert knowledge. From a computer scientist's viewpoint the project engaged many information integration challenges. I worked with Dr. Martin Doerr towards the *automatic place name identification problem*, a complex problem when integrating geo-spatial information. We developed a method that permits to estimate the precision of place name matching, the completeness of a gazetteer and the semantic inconsistency that relies in a digital gazetteer. Preliminary results have been presented in the NKOS Workshop of the European Conference on Digital Libraries [29]. Complete research results appeared in the IEEE Trans. on Knowl. and Data Engin. Journal (TKDE) [18]. In the scope of this project, I had the opportunity to become familiar with technologies for semantic web information integration, querying and dissemination.

3.2 Computer Supported Collaborative Work

Besides my interest in deploying quality research, I am very fascinated in building large-scale information systems with research extensions. Among others, such as the Movie Recommendation System dedicated to research purposes and an Online Questionnaire System, I would like to especially refer to "Confious" [21]. Confious[1] is a state-of-the-art conference management system that combines modern design, sophisticated algorithms and powerful engine so as to efficiently and professionally support the whole submission and reviewing process of an academic/research conference or workshop. Confious was initially selected by the University of Crete, Greece, as a student's innovative idea worth funding for further development. An industrial paper describing Confious was published at the WISE Conference [13], while it was also mentioned in the ERCIM News [20]. In the following years, Confious has successfully supported the submission and reviewing process of a number of International conferences and workshops, while significant is the interest of other committees for employing Confious for the organization of their upcoming conferences.

4 City Science and Urban Informatics

4.1 Online Socio-technical Analysis of Sustainable Buildings

The world is experiencing a period of extreme urbanization. Our need to improve our understanding of cities, however, is pressed not only by the social relevance of urban environments, but also by the availability of new strategies for city scale interventions that are enabled by emerging technologies. Designing and developing better tools and strategies for the creation of new cities, is therefore, a global imperative and defines new research initiatives (commonly referred to as: urban informatics, smart cities, city science, smart planet and so on) that try to leverage advances in data analysis, sensor technologies, and urban experiments to provide new insights into creating a data-driven approach to urban design and planning.

Towards this end, I have recently been involved in a research project named "Green2.0: A Middleware for Socio-technical Analytics of Green Buildings", lead by Prof. Tamer El-Diraby of the Civil Engineering Department at the University of Toronto. The project falls in the general research area of Smart Cities & Big Data and Analytics and its main objective is to design and develop an open middleware platform and web service that

[1] Confious conference management system. available online: http://www.confious.com.

enables: (a) online data management of Building Information Models (BIM), (b) online collaboration and interaction of professionals in the ACE (Architecture-Construction-Engineering) industry, (c) engagement of citizens and harnessing the public opinion in building early design stage to improve decision making and public policy, (d) sharing knowledge and educating the citizens of the critical trade-offs in building design (cost, functionality, energy efficiency, aesthetics) through what-if scenarios (e.g., gamification), (e) real-time energy efficiency analysis of buildings through integration with life-cycle database inventories and energy-efficiency simulation tools (EnergyPlus, OpenStudio, etc.), (f) social network analytics of the discussion networks around building elements, (g) the creation of a researcher's sandbox that offers RESTful APIs to (i) A BIM data repository, (ii) BIM Energy Efficiency Analysis repository, (iii) Communication-rich BIM-enabled social analytics repository. Preliminary results of this work have been published at the International Conference on Civil, Structural and Transportation Engineering (ICCSTE 2015) [1] and the International Conference on Smart Infrastructure and Construction (ICSIC 2016) [8]. Extensions of this work have been submitted as a demo paper [9], a full research paper to premier CS conference [7] and a top-tier AEC journal [19].

5 Statement

Throughout this document I have tried to establish a better understanding of my research interests and priorities. Discovering something new, gaining deep understanding of problems and solving them, being able to make an impact, doing things right, being the authority in an area and having my work cited by others are aspects of life that I would like to pursue and enjoy in my career. My short-term goal is to join a renowned academic or research institute, with world-class scientific staff, excellent resources, and industrial and academic partnerships that could help me grow as an educator and researcher. I am looking forward to a research environment that welcomes creativity, fosters the interchange of ideas and conversations, and encourages collaboration.

References

[1] **C** Alan Martin Redmond, M. Papagelis, and T. El-Diraby. Employing exploratory research stage to evaluate green building technologies for sustainable systems. In *International Conference on Civil, Structural and Transportation Engineering (ICCSTE)*, 2015.

[2] **C** Athanasios Papagelis, M. Papagelis, and C. D. Zaroliagis. Enabling social navigation on the web. In *Web Intelligence*, 2008.

[3] **C** Athanasios Papagelis, M. Papagelis, and C. D. Zaroliagis. Iclone: towards online social navigation. In *ACM Hypertext*, 2008.

[4] **C** Manos Papagelis. Refining social graph connectivity via shortcut edge addition. In *Hellenic Data Management Symposium.*, 2015.

[5] **C** Manos Papagelis, N. Bansal, and N. Koudas. Information cascades in the blogosphere: A look behind the curtain. In *ICWSM*, 2009.

[6] **C** Manos Papagelis, F. Bonchi, and A. Gionis. Suggesting ghost edges for a smaller world. In *ACM CIKM*, 2011.

[7] **C** Manos Papagelis, T. F. Krijnen, M. Elshenawy, and T. El-Diraby. Green2.0: An open platform for enabling online socio-technical analysis of sustainable buildings. In *WWW. Submitted*, 2016.

[8] **C** Manos Papagelis, T. F. Krijnen, M. Elshenawy, and T. El-Diraby. Green2.0: Towards real-time socio-technical analysis of building information models. In *International Conference on Smart Infrastructure and Construction (ICSIC)*, 2016.

[9] **C** Manos Papagelis, T. F. Krijnen, M. Elshenawy, T. Konomi, R. Fang, and T. El-Diraby. Green2.0: Enabling complex interactions between buildings and people. In *Computer-Supported Cooperative Work (ACM CSCW)*, 2016.

[10] **C** Manos Papagelis, V. Murdock, and R. van Zwol. Individual behavior and social influence in online social systems. In *ACM HyperText*, 2011.

[11] **C** Manos Papagelis and D. Plexousakis. Recommendation based discovery of dynamic virtual communities. In *CAiSE Short Paper Proceedings*, 2003.

[12] **C** Manos Papagelis, D. Plexousakis, and T. Kutsuras. Alleviating the sparsity problem of collaborative filtering using trust inferences. In *iTrust*, 2005.

[13] **C** Manos Papagelis, D. Plexousakis, and P. Nikolaou. Confious: Managing the electronic submission and reviewing process of scientific conferences. In *WISE*, 2005.

[14] **C** Manos Papagelis, I. Rousidis, D. Plexousakis, and E. Theoharopoulos. Incremental collaborative filtering for highly-scalable recommendation algorithms. In *ISMIS*, 2005.

[15] **J** Manos Papagelis. Refining social graph connectivity via shortcut edge addition. *ACM Trans. on Knowledge Discovery from Data (ACM TKDD)*, 10(2), 2015.

[16] **J** Manos Papagelis, G. Das, and N. Koudas. Sampling online social networks. *IEEE Trans. on Knowledge and Data Engineering (IEEE TKDE)*, 3(25), 2013.

[17] **J** Manos Papagelis and D. Plexousakis. Qualitative analysis of user-based and item-based prediction algorithms for recommendation agents. *Elsevier Eng. Applic. Artif. Intellig.*, 18(10), 2005.

[18] **J** Martin Doerr and M. Papagelis. A method for estimating the precision of placename matching. *IEEE Trans. on Knowledge and Data Engineering (IEEE TKDE)*, 19(8), 2007.

[19] **J** Tamer El-Diraby, T. F. Krijnen, and M. Papagelis. Green2.0: Integrated bim-based analysis of green buildings. *Automation in Construction. Submitted*, XX(X), 20XX.

[20] **M** Manos Papagelis and D. Plexousakis. Confious: Conference management system with intelligence, power and style. *ERCIM News*, 64(1), 2006.

[21] **O** Manos Papagelis. Confious conference management system. available online: http://www.confious.com.

[22] **P** Vanessa Murdock, R. V. Zwol, and E. Papangelis. Measuring or estimating user credibility. u.s. patent# 2012/0084226 a1., 04 2012.

[23] **P** Vanessa Murdock, R. V. Zwol, and E. Papangelis. Media or content tagging determined by user credibility signals. u.s. patent# 2012/0084302 a1., 04 2012.

[24] **T** Manos Papagelis. *Study and Implementation of User-based Recommendation Algorithms and their role in the Formation of Online Virtual Communities*. Bachelor's thesis, University of Toronto, Canada, 2003.

[25] **T** Manos Papagelis. Crawling the algorithmic foundations of recommendation technologies. Master's thesis, University of Crete, Greece, 2005.

[26] **T** Manos Papagelis. *Understanding and Improving Online Social Interactions and Processes: Methods, Algorithms & Applications*. PhD thesis, University of Toronto, Canada, 2015.

[27] **W** Gautam Das, N. Koudas, M. Papagelis, and S. Puttaswamy. Efficient sampling of information in social networks. In *ACM CIKM/SSM*, 2008.

[28] **W** Manos Papagelis and D. Plexousakis. Qualitative analysis of user-based and item-based prediction algorithms for recommendation agents. In *CIA*, 2004.

[29] **W** Martin Doerr and M. Papagelis. Precision estimation for matching place names to a digital gazetteer. In *ECDL/NKOS*, 2004.

Reference key:
C: Conference, **J**: Journal, **M**: Magazine, **O**: Other **P**: Patent, **T**: Thesis, **W**: Workshop