

# Homophily at Academic Conferences

Martin Atzmueller\*

Tilburg University, Department of Cognitive Science  
and Artificial Intelligence (CSAI)  
m.atzmuller@uvt.nl

Florian Lemmerich\*

RWTH Aachen University and  
GESIS - Leibniz Institute for the Social Sciences  
florian.lemmerich@humtec.rwth-aachen.de

## ABSTRACT

Academic conferences are a backbone for the exchange of ideas in scientific communities. However, so far little is known about the communication networks emerging at those venues. Besides personal knowledge, network homophily has been identified as a driving factor for establishing contacts and followerships in social networks, i. e., people are more likely to engage with others if they are similar with respect to certain attributes. In this paper, we describe work in progress on investigating homophily at four academic conferences based on face-to-face (F2F) contact data collected using wearable sensors between conference participants. In particular, we study which personal attributes are predictive for face-to-face contacts. For that purpose, we obtained diverse personal attributes from online sources in order to elicit a variety of hypotheses, which can then be compared using descriptive statistics and a Bayesian method for comparing hypotheses in networks. Our results suggest that personal knowledge (as derived from DBLP and ResearchGate networks) and homophilic behavior with respect to several attributes, e. g., gender or country of origin, are important factors for contacts at academic conferences.

## KEYWORDS

Homophily, Hypotheses, Network Analysis, Bayesian Statistics

### ACM Reference Format:

Martin Atzmueller\* and Florian Lemmerich\*. 2018. Homophily at Academic Conferences. In *WWW '18 Companion: The 2018 Web Conference Companion, April 23–27, 2018, Lyon, France*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3184558.3186953>

## 1 INTRODUCTION & RELATED WORK

Academic conferences are an integral part for forming scientific communities and serve the dissemination of research results, the exchange of ideas and the planning of future projects. They are a meeting point with acquaintances and (former) colleagues, but also bring together diverse researchers of different fields, origin, academic status or gender. Studying the types of contacts established at academic conferences would not only improve the understanding of scientific communities but could also encourage practical improvements for conference organization, e. g., for promoting diversity.

\* Both authors contributed equally.

This paper is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

*WWW '18 Companion, April 23–27, 2018, Lyon, France*

© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.

ACM ISBN 978-1-4503-5640-4/18/04.

<https://doi.org/10.1145/3184558.3186953>

For that purpose, this paper targets the investigation of social networks collected at four academic computer science conferences, i. e., Hypertext (HT) 2011, LWA 2010, LWA 2011, and LWA 2012. In particular, we consider face-to-face (F2F) contact data collected using wearable sensors<sup>1</sup> [2] between conference participants, in the context of the CONFERATOR system<sup>2</sup> [1, 4].

Our contributions are summarized as follows: (i) We investigate different (personal) attributes with respect to homophily in the context of contacts in academic conferences; we demonstrate their impact using descriptive statistics and Bayesian modeling. Our results indicate, that (ii) contacts to previous coauthors/formal friends are strongly predictive, but overall make up only a small part of the contact network; (iii) the network also exposes homophily in further different aspects, e. g., gender or country of origin.

## 2 DATASETS AND METHODS

Using the CONFERATOR system [1], we invited conference participants<sup>3</sup> to wear active RFID proximity tags. These can detect F2F contacts, i. e., close-range (about 1.5 meters) F2F proximity of the individuals wearing them, cf. [2]. This results in time-resolved networks of F2F contacts. Table 1 provides summary statistics of the collected datasets; for a detailed description, we refer to [4].

**Table 1: Summary statistics: Number of participants  $|V|$ , unique contacts  $|U|$ , average degree, diameter  $d$ , density, and count of F2F contacts  $|C|$ , cf. [4] for further details.**

Network	$ V $	$ U $	$\varnothing$ Degree	$d$	Density	$ C $
HT 2011	69	550	15.94	4	0.23	1902
LWA 2010	77	1004	26.08	3	0.34	5154
LWA 2011	42	300	14.29	3	0.35	1421
LWA 2012	44	354	16.09	3	0.37	1888

In addition to the F2F contacts of the participants, we obtained further information from their Conferator online profile and by crawling additional sources from the web. In particular, we utilize information on the participants' gender, country of origin, (university) affiliation, academic status (professor, postdoc, PhD, student), and their main conference track of interest. Note that not all attributes are available for all conferences (e. g., country is not available for LWA conferences since almost all participants were from Germany). Additionally, for each pair of participants we extracted friendship relations of the ResearchGate online platform<sup>4</sup>, and determined if they co-authored at least one paper together according to the DBLP publication database<sup>5</sup>.

<sup>1</sup><http://www.sociopatterns.org>

<sup>2</sup><http://www.conferator.org>

<sup>3</sup>Study participants also gave their informed consent for the use of their data (including their profile) in scientific studies.

<sup>4</sup><https://www.researchgate.net/>

<sup>5</sup><http://dblp.uni-trier.de/>

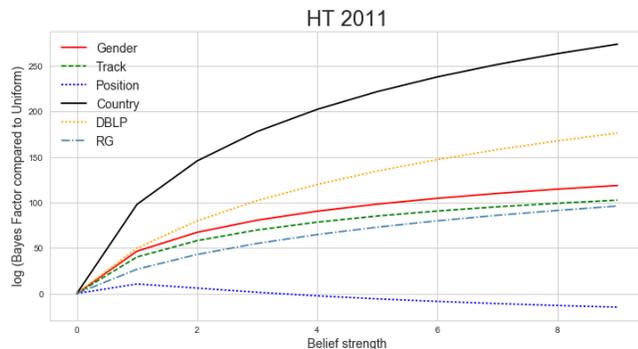
**Table 2: Homophily statistics for personal attributes: The first line shows the overall number of participant pairs, contacts, and average contacts per pair. Below, we compute for each attribute  $x$  the number of dyads (potential edges)  $E_x$  for participants with the same attribute value in  $x$ , the number of observed contacts  $C_x$  for these dyads, and the ratio between those two.**

(a) Results for HT 2011				(b) Results for LWA 2010				(c) Results for LWA 2011				(d) Results for LWA 2012			
Attribute $x$	$E_x$	$C_x$	$\frac{C_x}{E_x}$	Attribute $x$	$E_x$	$C_x$	$\frac{C_x}{E_x}$	Attribute $x$	$E_x$	$C_x$	$\frac{C_x}{E_x}$	Attribute $x$	$E_x$	$C_x$	$\frac{C_x}{E_x}$
Overall	2346	1902	0.81	Overall	2926	5154	1.76	Overall	861	1421	1.65	Overall	946	1888	2.00
Gender	1606	1613	1.00	Gender	2306	4501	1.95	Gender	645	1197	1.86	Gender	687	1418	2.06
Track	765	832	1.09	Track	1159	3080	2.66	Track	61	137	2.25	Track	490	1149	2.34
Position	763	697	0.91	Position	284	821	2.89	Position	269	453	1.68	Position	308	810	2.63
Country	198	529	2.67	Affiliation	53	484	9.13	Affiliation	27	252	9.33	Affiliation	49	572	11.67
ResearchGate	23	328	14.26	Researchgate	105	1130	10.76	ResearchGate	38	446	11.74	ResearchGate	35	332	9.49
DBLP	34	572	16.82	DBLP	86	1210	14.07	DBLP	68	794	11.68	DBLP	32	292	9.12

We analyze the data using descriptive statistics and the JANUS [3] method, a Bayesian approach for comparing hypotheses on network edge formation. In essence, JANUS transforms belief matrices, which the researcher specifies according to a hypothesis he has about the edges in a network, into Bayesian priors for a Dirichlet multinomial model for different strengths of beliefs in the hypothesis. Then, the marginal likelihood of the model given these priors can be compared against each other using Bayes Factors (roughly a Bayesian counterpart to a frequentists' p-value). In our setting, we construct hypotheses to be investigated as follows: For each personal attribute of a participant (e.g., gender, country, ...) we specify one homophily hypothesis that assumes that contacts between participants, for which this attribute is equal, are two times as likely. Furthermore, we specify hypotheses for the ResearchGate and DBLP networks, which express beliefs that a contact between two participant is twice as likely if they have a link in these networks. As a baseline, we use the uniform hypothesis, which states that all contacts are equally likely.

### 3 RESULTS AND DISCUSSION

Table 2 shows the descriptive results. It displays the number of pairs of participants that have the same value for attribute  $x$  (e.g., number of pairs for which either both are male or both are female), the overall number of face-to-face contacts between these pairs, and the average number of contacts per pair. Overall, we observe that the attributes *ResearchGate* and *DBLP*, which indicate a known personal



**Figure 1: Bayesian assessment of hypotheses for the HT 2011 conference. Higher y-values indicate higher plausibility of a hypothesis, higher x-values stronger belief in a hypothesis.**

relationship between individuals, lead to the highest increase of contact counts per pair. However, since these occur only seldomly, only a small fraction of the overall contacts can be explained by these. We also observe indicators for homophily in face-to-face contacts with respect to *country*, *track*, and *gender*, and most significantly *affiliation*. This suggests that participants are more likely to establish contacts that share personal attributes.

Figure 1 depicts an example result for comparing hypotheses. Each hypothesis is represented by one line. The x-axis shows different strengths of beliefs in a hypothesis, higher y-values indicate higher plausibility of the respective hypothesis in terms of a (log) Bayes Factor in comparison to the baseline uniform hypothesis. As we can see, the homophily hypothesis with respect to the attribute *country* has the highest plausibility for the Hypertext 2011 conference. This is due to a highly increased contact frequency between conference participants from the same country, while at the same time covering a decently sized share of all dyads. Homophily hypotheses regarding the attributes *gender*, and *track* are also significantly more plausible than a uniform hypothesis, in contrast to the homophilic hypothesis based on the academic *position*. The *DBLP* and *ResearchGate* hypotheses perform both significantly better than a uniform hypothesis. However, they cannot explain the data as well as the *country* hypothesis since they are too specific. These results are consistent with the descriptive analysis. Varying parameters in the hypotheses (e.g., assuming homophilic contacts to be three times as likely) lead to similar results.

Overall, our results indicate homophilic behavior of conference participants with respect to several personal attributes. In the future, we plan to extend the set of explored attributes and hypotheses, and to take network dynamics during the conference into account.

### REFERENCES

- [1] Martin Atzmueller, Dominik Benz, Stephan Doerfel, Andreas Hotho, Robert Jäschke, Bjoern Elmar Macek, Folke Mitzlaff, Christoph Scholz, and Gerd Stumme. 2011. Enhancing Social Interactions at Conferences. *it - Information Technology* 53, 3 (2011), 101–107.
- [2] Alain Barrat, Ciro Cattuto, Vittoria Colizza, Jean-François Pinton, Wouter Van den Broeck, and Alessandro Vespignani. 2008. High Resolution Dynamical Mapping of Social Interactions With Active RFID. *ArXiv e-prints* (Nov. 2008). arXiv:cs.CY/0811.4170
- [3] Lisette Espin-Noboa, Florian Lemmerich, Markus Strohmaier, and Philipp Singer. 2017. JANUS: A hypothesis-driven Bayesian approach for understanding edge formation in attributed multigraphs. *Applied Network Science* 2, 1 (2017), 16.
- [4] Mark Kibanov, Martin Atzmueller, Jens Illig, Christoph Scholz, Alain Barrat, Ciro Cattuto, and Gerd Stumme. 2015. Is Web Content a Good Proxy for Real-Life Interaction? A Case Study Considering Online and Offline Interactions of Computer Scientists. In *Proc. ASONAM*. IEEE Press, Boston, MA, USA.