Haewoon Kwak Curriculum vitae

CONTACT Qatar Computing Research Institute

Hamad Bin Khalifa University

Mobile: +974-6606-3447

Doha, Qatar

Web: http://haewoon.io

E-mail: haewoon@acm.org

HIGHLIGHTS

Highly cited research (H-index: 22, i10-index: 35)

- Twitter as an information network [C8]: 7,265 times
- YouTube video popularity characterization [C3]: 1,816 times
- Huge online social network (OSN) analysis [C2]: 1,196 times

Computational journalism

- Media attention and public attention: The Web'18 [C35]
- Media attention and disregard network: ASONAM'17 [C32]
- Sentiment of news headlines: ICWSM'15 [C22] (featured by O Globo)
- Bias in foreign news coverage: Computation+Journalism Symposium'14 (featured by MIT Technology Review, ACM Tech News, etc.)
- Tanbih news aggregator: EMNLP'19 Demo [C45] (featured by Forbes, MIT Technology Review, Wired, Engadget, etc.)

Computational Social Science

- Polarization on social media: ICWSM'19 [C42]; ICWSM'16 [C23]
- User behavior in extreme situations: WWW Companion'17 [C27] (covered by New Scientist, Daily Mail, etc.)
- Cyberbullying and toxic behavior and in online games: CHI'15 [C21]; WWW'14 [C15] (featured by Nature News, Scientific American, etc.)
- Alt-right community analysis: WWW Companion'18 [C38] (featured by New Scientist, Vice, etc.)
- Structures and dynamics of social media: WWW'10 [C8] (featured by Mashable Op-Ed, Readwrite, The Guardian, MIT Technology Review, etc.); CHI'11 [C10] (featured by Kyunghyang); ICWSM'12 [C12]; CSCW'13 [C13]

Interests

My research interests are data-driven modeling and large-scale analysis of online systems from social media to online games to better understand social phenomena. Recently, I more focus on understanding the obstacles to the trusted public space online in a data-driven way. In particular, I define *media bias*, *toxicity*, *polarization*, and *unfair representations* as the four major obstacles to the trusted web. I aim to develop methodologies to model them, construct transparent frameworks to monitor them at large-scale in real-time, and make the online public space more trustable.

EDUCATION

Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea

Ph.D. in Computer Science

August 2011

Advisor: Professor Sue Moon

Dissertation: Multi-level Analysis on Structures and Dynamics of Online Social

Networks

M.S. in Computer Science

August 2007

B.S. in Computer Science

Feburary 2006

Professional Experience

Qatar Computing Research Institute, HBKU, Doha, Qatar

Senior Scientist Jul 2016 - Present

Scientist Apr 2014 - Jun 2016

Telefonica Research, Barcelona, Spain

Researcher Jun 2012 - Apr 2014

KAIST, Daejeon, Korea

Research Assistant Aug 2011 - Oct 2011

Telefonica Research, Barcelona, Spain

Intern Sep 2007 - Dec 2007

AWARDS

- 2019 Best paper in International Conference on Social Informatics (SocInfo) '19
- 2018 HBKU Innovation Idea Development Grants, Doha, Qatar
- 2016 Qatar Science & Technology Park (QSTP) Accelerator program, Doha, Qatar
- 2011 Best Ph.D. Thesis in KAIST Computer Science Department, Daejeon, Korea
- 2007 Best paper award in ACM Internet Measurement Conference '17

Publications Journals

- [J7] A. Zubiaga, B. Heravi, J. An, and H. Kwak. Social media mining for journalism. Online Information Review, 43(1):2–6, 2019
- [J6] J. An, H. Kwak, S. Jung, J. Salminen, M. Admad, and B.J. Jansen. Imaginary people representing real numbers: generating personas from online social media data. ACM Transactions on the Web (TWEB), 12(4):27, 2018
- [J5] P. Mawhorter, S. Şengün, H. Kwak, and D.F. Harrell. Identifying regional trends in avatar customization. *IEEE Transactions on Games*, pages 1–11, 2018
- [J4] S.-Y. Shin, D.-W. Seo, J. An, H. Kwak, S.-H. Kim, J. Gwack, and M.-W. Jo. High correlation of Middle East respiratory syndrome spread with Google search and Twitter trends in Korea. Scientific reports, 6:32920, 2016
- [J3] H. Park, H. Kwak, M. Cha, and S. Moon. Influentials ranking in social networks. Communications of the Korean Institute of Information Scientists and Engineers (in Korean), 28(3):24–30, 2010
- [J2] H. Kwak, S. Moon, Y. Eom, Y. Choi, and H. Jeong. Consistent community identification in complex networks. *Journal of the Korean Physical Society*, 59(5):3128–3132, 2011
- [J1] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon. Analyzing the video popularity characteristics of large-scale user generated content systems. *IEEE/ACM Transactions on Networking*, 17(5):1357–1370, Oct 2009

Conferences

- [C45] Y. Zhang, G.D.S. Martino, A. Barrón-Cedeño, S. Romeo, J. An, H. Kwak, T. Staykovski, I. Jaradat, G. Karadzhov, R. Baly, K. Darwish, J. Glass, and P. Nakov. Tanbih: Get to know what you are reading. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations, pages 223–228, Hong Kong, China, November 2019. Association for Computational Linguistics
- [C44] J. An and H. Kwak. Gender and racial diversity in commercial brands' advertising images on social media. In *International Conference on Social Informatics (SocInfo)*. Springer, 2019
- [C43] H. Almerekhi, H. Kwak, B.J. Jansen, and J. Salminen. Detecting toxicity triggers in online discussions. In *Proceedings of the 30th ACM Conference on Hypertext* and Social Media, pages 291–292. ACM, 2019
- [C42] J. An, H. Kwak, O. Posegga, and A. Jungherr. Political discussions in homogeneous and cross-cutting communication spaces. In *Proceedings of the International AAAI* Conference on Web and Social Media (ICWSM), volume 13, pages 68–79, 2019
- [C41] S. Jung, J. An, H. Kwak, J. Salminen, and B.J. Jansen. Assessing the accuracy of four popular face recognition tools for inferring gender, age, and race. In *Twelfth International AAAI Conference on Web and Social Media (ICWSM)*, pages 624–627, 2018
- [C40] J. Salminen, H. Almerekhi, M. Milenković, S. Jung, J. An, H. Kwak, , and B.J. Jansen. Anatomy of online hate: developing a taxonomy and machine learning models for identifying and classifying hate in online news media. In Twelfth International AAAI Conference on Web and Social Media, pages 330–339, 2018
- [C39] J. An, H. Kwak, and Y.-Y. Ahn. SemAxis: A lightweight framework to characterize domain-specific word semantics beyond sentiment. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 2450–2461, Melbourne, Australia, July 2018. Association for Computational Linguistics
- [C38] S. Zannettou, B. Bradlyn, E. De Cristofaro, H. Kwak, M. Sirivianos, G. Stringini, and J. Blackburn. What is Gab: A bastion of free speech or an alt-right echo chamber. In Companion Proceedings of the The Web Conference 2018, pages 1007–1014. International World Wide Web Conferences Steering Committee, 2018
- [C37] J. Salminen, B.J. Jansen, J. An, S. Jung, L. Nielsen, and H. Kwak. Fixation and confusion: Investigating eye-tracking participants' exposure to information in personas. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, pages 110–119. ACM, 2018
- [C36] J. Salminen, H. Kwak, J. Santos, S. Jung, J. An, and B.J. Jansen. Persona perception scale: developing and validating an instrument for human-like representations of data. In Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, page LBW075. ACM, 2018
- [C35] H. Kwak, J. An, J. Salminen, S.-G. Jung, and B. Jansen. What we read, what we search: Media attention and public attention among 193 countries. In *Proceedings of the 2018 World Wide Web Conference*, pages 893–902. International World Wide Web Conferences Steering Committee, 2018
- [C34] J. An and H. Kwak. Multidimensional analysis of the news consumption of different demographic groups on a nationwide scale. In *International Conference* on Social Informatics (SocInfo), pages 124–142. Springer, 2017
- [C33] J. An, H. Aldarbesti, and H. Kwak. Convergence of media attention across 129 countries. In *International Conference on Social Informatics (SocInfo)*, pages 159– 168. Springer, 2017

- [C32] H. Kwak and J. An. Multiplex media attention and disregard network among 129 countries. In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) 2017, pages 1225–1232. ACM, 2017
- [C31] J. Reis, H. Kwak, J. An, J. Messias, and F. Benevenuto. Demographics of news sharing in the US Twittersphere. In Proceedings of the 28th ACM Conference on Hypertext and Social Media, pages 195–204. ACM, 2017
- [C30] D.F. Harrell, S. Vieweg, H. Kwak, C.-U. Lim, S. Sengun, A. Jahanian, and P. Ortiz. Culturally-grounded analysis of everyday creativity in social media: A case study in qatari context. In *Proceedings of the 2017 ACM SIGCHI Conference* on Creativity and Cognition, pages 209–221. ACM, 2017
- [C29] J. An and H. Kwak. Data-driven approach to measuring the level of press freedom using media attention diversity from unfiltered news. In The Workshops of News and Public Opinion (NECO) co-located with Eleventh International AAAI Conference on Web and Social Media (AAAI Technical Report WS-17-17), 2017
- [C28] J. An and H. Kwak. What gets media attention and how media attention evolves over time: large-scale empirical evidence from 196 countries. In *Eleventh International AAAI Conference on Web and Social Media (ICWSM)*, pages 464–467, 2017
- [C27] A.R. Kang, J. Blackburn, H. Kwak, and H.K. Kim. I would not plant apple trees if the world will be wiped: Analyzing hundreds of millions of behavioral records of players during an MMORPG beta test. In *Proceedings of the 26th International* Conference on World Wide Web Companion, pages 435–444. International World Wide Web Conferences Steering Committee, 2017
- [C26] K. Park, M. Cha, H. Kwak, and K.-T. Chen. Achievement and friends: Key factors of player retention vary across player levels in online multiplayer games. In Proceedings of the 26th International Conference on World Wide Web Companion, pages 445–453. International World Wide Web Conferences Steering Committee, 2017
- [C25] H. Kwak and J. An. Revealing the hidden patterns of news photos: Analysis of millions of news photos through GDELT and deep learning-based vision APIs. In The workshop on News and Public Opinion co-located with Tenth International AAAI Conference on Web and Social Media (AAAI Technical Report WS-16-18), 2016
- [C24] H. Kwak and J. An. Two tales of the world: Comparison of widely used world news datasets GDELT and EventRegistry. In *Tenth International AAAI Conference on Web and Social Media*, pages 619–622, 2016
- [C23] J. An, H. Kwak, Y. Mejova, S.A.S. De Ogar, and B.G. Fortes. Are you Charlie or Ahmed? cultural pluralism in Charlie Hebdo response on Twitter. In *Tenth International AAAI Conference on Web and Social Media (ICWSM)*, pages 2–11, 2016
- [C22] D. Reis, F. Benevenuto, P. Olmo, P. Raquel, H. Kwak, and J. An. Breaking the news: First impressions matter on online news. In Ninth International AAAI Conference on Web and Social Media (ICWSM), 2015
- [C21] H. Kwak, J. Blackburn, and S. Han. Exploring cyberbullying and other toxic behavior in team competition online games. In *Proceedings of the 33rd Annual* ACM Conference on Human Factors in Computing Systems (CHI), pages 3739– 3748. ACM, 2015
- [C20] I. Leontiadis, A. Lima, H. Kwak, R. Stanojevic, D. Wetherall, and D. Papagiannaki. From cells to streets: Estimating mobile paths with cellular-side data. In *Proceedings* of the 10th ACM International on Conference on emerging Networking Experiments and Technologies (CoNEXT), pages 121–132. ACM, 2014

- [C19] H. Kwak and J. An. A first look at global news coverage of disasters by using the GDELT dataset. In *International Conference on Social Informatics (SocInfo)*, pages 300–308. Springer, 2014
- [C18] H. Kwak and J. Blackburn. Linguistic analysis of toxic behavior in an online video game. In Workshop on Exploration on Games and Gamers co-located with international Conference on Social Informatics, pages 209–217. Springer, 2014
- [C17] L. Gyarmati, H. Kwak, and P. Rodriguez. Searching for a unique style in soccer. In 2014 KDD Workshop on Large-Scale Sports Analytics, 2014
- [C16] M. Martin, R.D. Oliveira, H. Kwak, , and N. Oliver. Didn't you see my message?: predicting attentiveness to mobile instant messages. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3319–3328. ACM, 2014
- [C15] J. Blackburn and H. Kwak. STFU NOOB!: predicting crowdsourced decisions on toxic behavior in online games. In *Proceedings of the 23rd international conference* on World wide web, pages 877–888. ACM, 2014
- [C14] Y. Hong, H. Kwak, Y. Baek, and M. Sue. Tower of babel: A crowdsourcing game building sentiment lexicons for resource-scarce languages. In *Proceedings of* the 22nd International Conference on World Wide Web, WWW '13 Companion, pages 549–556, New York, NY, USA, 2013. ACM
- [C13] B. Xu, Y. Huang, H. Kwak, and N. Contractor. Structures of broken ties: Exploring unfollow behavior on Twitter. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, CSCW '13, pages 871–876, New York, NY, USA, 2013. ACM
- [C12] H. Kwak, S. Moon, and W. Lee. More of a receiver than a giver: why do people unfollow in Twitter? In Sixth International AAAI Conference on Weblogs and Social Media, pages 499–502, 2012
- [C11] H. Kwak, Y. Hong, J. You, and S. Moon. Visualizing a personal timeline by adding multiple social contexts. In The workshop on Social Media Visualization co-located with International AAAI Conference on Web and Social Media (AAAI Technical Report WS-12-03), 2012
- [C10] H. Kwak, H. Chun, and S. Moon. Fragile online relationship: a first look at unfollow dynamics in Twitter. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 1091–1100. ACM, 2011
- [C9] C. Lee, H. Kwak, H. Park, and S. Moon. Finding influentials based on the temporal order of information adoption in Twitter. In *Proceedings of the 19th* international conference on World wide web, pages 1137–1138. ACM, 2010
- [C8] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? In Proceedings of the 19th international conference on World wide web, pages 591–600. ACM, 2010
- [C7] S. Moon, J. You, H. Kwak, D. Kim, and H. Jeong. Understanding topological mesoscale features in community mining. In 2010 Second International Conference on COMmunication Systems and NETworks (COMSNETS 2010), pages 1–10. IEEE, 2010
- [C6] H. Kwak, Y. Choi, Y. Eom, H. Jeong, and S. Moon. Mining communities in networks: a solution for consistency and its evaluation. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement*, pages 301–314. ACM, 2009
- [C5] X. Amatriain, N. Lathia, J. Puyol, H. Kwak, and N. Oliver. The wisdom of the few: a collaborative filtering approach based on expert opinions from the web. In Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, pages 532–539. ACM, 2009

- [C4] H. Chun, H. Kwak, Y. Eom, Y.-Y. Ahn, S. Moon, and H. Jeong. Comparison of online social relations in volume vs interaction: A case study of cyworld. In Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement, IMC '08, pages 57–70, New York, NY, USA, 2008. ACM
- [C3] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon. I tube, you tube, everybody tubes: Analyzing the world's largest user generated content video system. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, IMC '07, pages 1–14, New York, NY, USA, 2007. ACM
- [C2] Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong. Analysis of topological characteristics of huge online social networking services. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, pages 835–844, New York, NY, USA, 2007. ACM
- [C1] H. Kwak, S. Han, Y.-Y. Ahn, S. Moon, and H. Jeong. Impact of snowball sampling ratios on network characteristics estimation: A case study of cyworld. In Proceedings of the Korean Information Science Society Conference, 2006

Patents

- [P4] N. Oliver M. Pielot, H. Kwak, and R. Oliveira. Method for predicting reactiveness of users of mobile devices for mobile messaging, June 25 2015. US Patent App. 14/577,488
- [P3] **H. Kwak** and E. Martinez. Method, system and computer-readable storage mediums for estimating a route, January 8 2015. US Patent App. 14/323,186
- [P2] S. Moon, H. Kwak, H. Shin, and J. Yoon. System and method for extracting users of similar interests between various types of web servers, April 16 2013. US Patent 8,423,542
- [P1] X. Amatriain, J. Puyol, N. Lathia, P. Rodriguez, and H. Kwak. Recommender system based on expert opinions, June 9 2011. US Patent App. 12/630,606

Advising Hamad Bin Khalifa University, Doha, Qatar

 ${\it Co-supervisor} \\ {\it Hind Al-merekhi}, PhD candidate at Hamad Bin Khalifa University} \\$

Professional Service

Editorial services

- Editorial board member, Online Social Networks and Media (Science Citation Index Expanded journal): 2019-present
- Editorial board member, Digital Journalism (2018 Impact Factor 2.679): 2020-

Organization

- Co-organizer, Workshop on Bias, Disinformation, Misinformation, and Propaganda in Online News and in Social Media co-located with SocInfo'19, Doha, Qatar (2019)
- Workshop chair, 29th ACM Conference on Hypertext and Social Media (HT), Baltimore, Maryland, USA (2018)
- Co-organizer, 3rd International Workshop on NEws and public Opinion (NECO 2018) co-located with ICWSM'18, Stanford, CA, USA (2018)
- Co-organizer, 4th workshop on Social News on the Web (SNOW 2016) Co-located with WWW'17, Perth, Australia (2017)

- Co-organizer, 2nd International Workshop on NEws and public Opinion (NECO 2017) co-located with ICWSM'17, Montreal, Canada (2017)
- Co-organizer, 3rd workshop on Social News on the Web (SNOW 2016) Co-located with WWW'16, Montreal, Canada (2016)
- Co-organizer, 1st International Workshop on NEws and public Opinion (NECO 2016) co-located with ICWSM'16, Cologne, Germany (2016)
- Track chair (with Alessandra Sala and Aristides Gionis), Social Networks and Graph Analysis Track in 2016 International World Wide Web Conference, Montreal, Canada (2016)
- Co-organizer, 1st International Workshop on Exploration Games and Gamers (EGG) co-located with SocInfo'14, Barcelona, Spain (2014)
- Registration chair, ACM Internet Measurement Conference (IMC), Barcelona, Spain (2013)

Program Committee Member

- The Web (WWW) Web and Society Track: 2018, 2019, 2020
- The Web (WWW) Social Network and Graph Analysis: 2014, 2015, 2016, 2018, 2019
- The Web (WWW) Poster: 2014
- ICWSM Senior PC: 2014, 2018, 2019, 2020
- ICWSM PC: 2013
- WWW Cybersafety Workshop: 2018, 2019
- ACL Abusive Language Workshop (ALW): 2019
- CKIM Rumours and Deception in Social Media Workshop: 2018
- Complex Networks: 2018
- IEEE ICDCS Social Networks and Crowdsourcing: 2018
- European Symposium Societal Challenges in Computational Social Science: 2017
- SocInfo Senior PC: 2014, 2015, 2016
- LC2S2 (At the crossroads: lessons and challenges in Computational Social Science): 2016
- ACL Workshop on NLP and Social Dynamics: 2014
- WebSci: 2014
- WWW Connecting online and offline life (COOL) Workshop: 2014
- \bullet WWW Simplifying Complex Networks for Practitioners (SIMPLEX) Workshop: 2014
- SocInfo Interaction and Exchange in Social Media Workshop: 2014
- ACM Hypertext: 2014
- European Conference on Complex Systems (ECCS): 2013
- CIKM Search and Mining User-generated Contents Workshop: 2013

IN THE PRESS Selected list of media coverages

- "What is Gab? A Bastion of Free Speech or an Alt-Right Echo Chamber?" (WWW Cybersafety workshop)
 - New Scientist
 - Vice
- "Data-driven Approach to Measuring the Level of Press Freedom Using Media Attention Diversity from Unfiltered News" (ICWSM'17)
 - The Best of the Physics arXiv (week ending April 15, 2017) in MIT Technology Review
- "I Would Not Plant Apple Trees If the World Will Be Wiped: Analyzing Hundreds of Millions of Behavioral Records of Players During an MMORPG Beta Test" (WWW Companion'17)
 - New Scientist
 - Daily Mail
 - IFL Science (18.1K shared)
 - PC Gamer
 - Massively Overpowered
 - 4Gamers
 - El Confidencial
 - Joongang Ilbo
 - How Stuff Works
 - PC Gamer
 - Digital Trends
 - Engadget
- "Revealing the Hidden Patterns of News Photos: Analysis of Millions of News Photos Using GDELT and Deep Learning-based Vision APIs" (ICWSM Workshop on News and Public Opinion'16)
 - The Best of the Physics arXiv (week ending March 26, 2016) in MIT Technology Review
- "Breaking the News: First Impressions Matter on Online News" (ICWSM'15)
 - Other Interesting arXiv Papers (Week ending April 11, 2015) in MIT Technology Review
 - O Globo
- "Understanding News Geography and Major Determinants of Global News Coverage of Disasters" (Computation+Journalism Symposium'14)
 - MIT Technology Review
 - ACM Tech News
- "Searching for a Unique Style in Soccer" (KDD Workshop on Large-scale Sports Analytics'14)
 - BBC
 - MIT Technology Review
 - The Sunday Times
 - The Economist
 - Slate

- Pacific Standard
- "Linguistic Analysis of Toxic Behavior in an Online Video Game" (SocInfo Workshop on Exploration Games and Gamers'14)
 - Other Interesting arXiv Papers (Week ending November 1, 2014) in MIT Technology Review
- "STFU NOOB! Predicting Crowdsourced Decisions on Toxic Behavior in Online Games" (WWW'14)
 - Nature News
 - Scientific American
 - Chosun Ilbo
- "Fragile Online Relationship: a First Look at Unfollow Dynamics in Twitter" (CHI'11)
 - Kyunghyang Shinmun
- "What is Twitter, a Social Network or a News Media?" (WWW'10)
 - Mashable
 - Readwrite
 - Giagom
 - The Guardian
 - PC World
 - MIT Technology Review
 - Chosun Ilbo
 - Dong-A Ilbo

References

Prof. Sue Moon

Professor, School of Computing, KAIST, Korea sbmoon@kaist.edu

Prof. Yong-Yeol Ahn

Associate Professor, School of Informatics, Computing, and Engineering, Indiana University Bloomington, USA yyahn@iu.edu

Prof. Meeyoung Cha

Associate Professor, School of Computing, KAIST, Korea Chief Investigator, Institute for Basic Science (IBS), Korea meeyoungcha@kaist.ac.kr