

of unprotected sex requests or methamphetamine use during encounters in Los Angeles? What is the distribution of those behaviors in ads targeting West Hollywood? Typically these types of questions are answered through survey data. Manually identifying search terms associated with high-risk behaviors (e.g., “bareback” as slang for unprotected sex, “my friend tina” as code for methamphetamine”, etc.) and geographically indexing ad text provides an inexpensive, automated way of augmenting survey data. We leave these ideas for future exploration.

REFERENCES

- [1] Centers for Disease Control and Prevention (CDC). (2014, Feb) HIV among African Americans. [Online]. Available: http://www.cdc.gov/hiv/pdf/risk_HIV_AfricanAmericans.pdf
- [2] H. Jaffe, R. Valdiserri, and K. De Cock, “The reemerging HIV/AIDS epidemic in men who have sex with men,” *JAMA: the journal of the American Medical Association*, vol. 298, no. 20, pp. 2412–2414, 2007.
- [3] T. Sanchez, T. Finlayson, A. Drake, S. Behel, M. Cribbin, E. DiNenno, T. Hall, S. Kramer, A. Lansky, C. for Disease Control, and P. (US), *Human Immunodeficiency Virus (HIV) Risk, Prevention, and Testing Behaviors: United States, National HIV Behavioral Surveillance System: Men who Have Sex Men, November 2003 [to] April 2005*. US Department of Health and Human Services, 2006.
- [4] D. Osmond, L. Pollack, J. Paul, and J. Catania, “Changes in prevalence of HIV infection and sexual risk behavior in men who have sex with men in San Francisco: 1997–2002,” *Journal Information*, vol. 97, no. 9, 2007.
- [5] M. McFarlane, S. Bull, and C. Rietmeijer, “The internet as a newly emerging risk environment for sexually transmitted diseases,” *JAMA: the journal of the American Medical Association*, vol. 284, no. 4, pp. 443–446, 2000.
- [6] A. Liao, G. Millett, and G. Marks, “Meta-analytic examination of online sex-seeking and sexual risk behavior among men who have sex with men,” *Sexually transmitted diseases*, vol. 33, no. 9, p. 576, 2006.
- [7] B. Rosser, W. West, and R. Weinmeyer, “Are gay communities dying or just in transition? Results from an international consultation examining possible structural change in gay communities,” *AIDS care*, vol. 20, no. 5, pp. 588–595, 2008.
- [8] E. Benotsch, S. Kalichman, and M. Cage, “Men who have met sex partners via the internet: Prevalence, predictors, and implications for HIV prevention,” *Archives of Sexual Behavior*, vol. 31, no. 2, pp. 177–183, 2002.
- [9] Alexa: The Web Information Company. (2013, September) Craigslist.org site info. [Online]. Available: <http://www.alexa.com/siteinfo/craigslist.org>
- [10] S. Wakamiya, R. Lee, and K. Sumiya, “Crowd-sourced urban life monitoring: Urban area characterization based crowd behavioral patterns from Twitter,” in *Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication*. ACM, 2012, p. 26.
- [11] K. M. Blankenship, S. J. Bray, and M. H. Merson, “Structural interventions in public health,” *Aids*, vol. 14, pp. S11–S21, 2000.
- [12] J. Kraut-Becher, M. Eisenberg, C. Voytek, T. Brown, D. S. Metzger, and S. Aral, “Examining racial disparities in HIV: Lessons from sexually transmitted infections research,” *JAIDS Journal of Acquired Immune Deficiency Syndromes*, vol. 47, pp. S20–S27, 2008.
- [13] K. Bernstein, F. Curriero, J. Jennings, G. Olthoff, E. Erbeling, and J. Zenilman, “Defining core gonorrhea transmission utilizing spatial data,” *American journal of epidemiology*, vol. 160, no. 1, pp. 51–58, 2004.
- [14] C. Fichtenberg and J. Ellen, “Moving from core groups to risk spaces,” *Sexually Transmitted Diseases*, vol. 30, no. 11, p. 825, 2003.
- [15] J. Chan and A. Ghose, “Internet’s dirty secret: Assessing the impact of technology shocks on the outbreaks of sexually transmitted diseases,” 2012.
- [16] California Department of Public Health. (2011, Dec) California syphilis elimination surveillance data. [Online]. Available: <http://www.cdph.ca.gov/data/statistics/Documents/STD-Data-Syphilis-Elimination-Surveillance-Data.pdf>
- [17] D. A. Moskowitz and D. W. Seal, ““GWM looking for sex-SERIOUS ONLY”: The interplay of sexual ad placement frequency and success on the sexual health of “men seeking men” on Craigslist,” *Journal of Gay & Lesbian Social Services*, vol. 22, no. 4, pp. 399–412, 2010.
- [18] C. Grov, “Risky sex-and drug-seeking in a probability sample of men-for-men online bulletin board postings,” *AIDS and Behavior*, vol. 14, no. 6, pp. 1387–1392, 2010.
- [19] J. A. Fries, A. M. Segre, and P. M. Polgreen, “Using online classified ads to identify the geographic footprints of anonymous, casual sex-seeking individuals,” in *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom)*. IEEE, 2012, pp. 402–410.
- [20] —, “Towards linking anonymous authorship in casual sexual encounter ads,” *Online Journal of Public Health Informatics*, vol. 5, no. 1, 2013.
- [21] J. A. Fries, A. M. Segre, L. Polgreen, and P. M. Polgreen, “The use of Craigslist posts for risk behavior and STI surveillance,” *International Society for Disease Surveillance Conference 2010*, p. 13, 2011.
- [22] J. A. Fries, A. Ho, A. M. Segre, and P. M. Polgreen, “Using Craigslist messages for syphilis surveillance,” in *International Meeting on Emerging Diseases and Surveillance (IMED)*, 2011.
- [23] S. B. Johnson and C. Friedman, “Integrating data from natural language processing into a clinical information system,” in *Proceedings of the AMIA Annual Fall Symposium*. American Medical Informatics Association, 1996, p. 537.
- [24] A. Mislove, S. Lehmann, Y.-Y. Ahn, J.-P. Onnela, and J. N. Rosenquist, “Understanding the demographics of Twitter users,” *ICWSM*, vol. 11, p. 5th, 2011.
- [25] K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis, “Tastes, ties, and time: A new social network dataset using Facebook.com,” *Social Networks*, vol. 30, no. 4, pp. 330–342, 2008.
- [26] R. Bhopal, “Glossary of terms relating to ethnicity and race: for reflection and debate,” *Journal of Epidemiology and Community Health*, vol. 58, no. 6, pp. 441–445, 2004.
- [27] U. Boehmer, N. R. Kressin, D. R. Berlowitz, C. L. Christiansen, L. E. Kazis, and J. A. Jones, “Self-reported vs administrative race/ethnicity data and study results,” *American Journal of Public Health*, vol. 92, no. 9, pp. 1471–1472, 2002.
- [28] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python*. O’Reilly Media, 2009.
- [29] J. A. Fries and A. M. Segre, “Geographic entity linking in online classified ads,” 2014, in-review.
- [30] United States Census Bureau. (2013, August) 2000 redistricting data (public law 94-171) summary file. [Online]. Available: <http://www.census.gov/prod/cen2000/doc/p194-171.pdf>
- [31] J. Patrick and M. Li, “High accuracy information extraction of medication information from clinical notes: 2009 i2b2 medication extraction challenge,” *Journal of the American Medical Informatics Association*, vol. 17, no. 5, pp. 524–527, 2010.
- [32] J. Lafferty, A. McCallum, and F. C. Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data,” 2001.
- [33] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2011, ISBN 3-900051-07-0. [Online]. Available: <http://www.R-project.org/>
- [34] A. E. Magurran, “Measuring biological diversity,” 2004.
- [35] P. F. Brown, P. V. Desouza, R. L. Mercer, V. J. D. Pietra, and J. C. Lai, “Class-based n-gram models of natural language,” *Computational Linguistics*, vol. 18, no. 4, pp. 467–479, 1992.
- [36] S. Miller, J. Guinness, and A. Zamanian, “Name tagging with word clusters and discriminative training,” in *HLT-NAACL*, vol. 4. Citeseer, 2004, pp. 337–342.