[32] Twitter. Twitter reporting impersonation accounts, 2014. https://support.twitter.com/articles/20170142-reporting-impersonation-accounts.

[33] B. Viswanath, M. A. Bashir, M. Crovella, S. Guha, K. Gummadi, B. Krishnamurthy, and A. Mislove. Towards detecting anomalous user behavior in online social networks. In *USENIX Security'14*.

[34] B. Viswanath, M. A. Bashir, M. B. Zafar, L. Espin, K. P. Gummadi, and A. Mislove. Trulyfollowing: Discover twitter accounts with suspicious followers. http://trulyfollowing.app-ns.mpi-sws.org/, April 2012. Last accessed Sept 6, 2015.

[35] B. Viswanath, M. Mondal, A. Clement, P. Druschel, K. Gummadi, A. Mislove, and A. Post. Exploring the design space of social network-based sybil defenses. In *COMSNETS'12*.

[36] B. Viswanath, A. Post, K. P. Gummadi, and A. Mislove. An analysis of social network-based sybil defenses. In *SIGCOMM '10*.

[37] G. Wang, M. Mohanlal, C. Wilson, X. Wang, M. J. Metzger, H. Zheng, and B. Y. Zhao. Social turing tests: Crowdsourcing sybil detection. In *NDSS'13*.

[38] Wikibin. Employers using social networks for screening applicants, 2008. http://wikibin.org/articles/employers-using-social-networks-for-screening-applicants.html.

[39] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman. Sybilguard: Defending against sybil attacks via social networks. In *SIGCOMM '06*.

[40] C. M. Zhang and V. Paxson. Detecting and analyzing automated activity on twitter. In *PAM'11*.

APPENDIX

Here we first explain how we computed similarity between various attribute values (e.g., names and photos) of accounts and then describe the procedure we used to determine when two attribute values (e.g., two names or two photos) are "similar enough" to be deemed to represent the same entity.

A. SIMILARITY METRICS

Name similarity Previous work in the record linkage community showed that the *Jaro string distance* is the most suitable metric to compare similarity between names both in the offline and online worlds [7, 23]. So we use the Jaro distance to measure the similarity between user-names and screen-names.

Photo similarity Estimating photo similarity is tricky as the same photo can come in different formats. To measure the similarity of two photos while accounting for image transformations, we use two matching techniques: (*i*) *perceptual hashing*, a technique originally invented for identifying illegal copies of copyrighted content that works by reducing the image to a transformation-resilient "fingerprint" containing its salient characteristics [24] and (*ii*) *SIFT*, a size invariant algorithm that detects local features in an image and checks if two images are similar by counting the number of local features that match between two images [18]. We use two different algorithms for robustness. The perceptual hashing technique does not cope well with some images that are resized, while the SIFT algorithm does not cope well with computer generated images.

Location similarity For all profiles, we have the textual representations of the location, like the name of a city. Since social networks use different formats for this information, a simple textual comparison will be inaccurate. Instead, we convert the location to latitude/longitude coordinates by submitting them to the Bing API [1]. We then compute the similarity between two locations as the actual geodesic distance between the corresponding coordinates.

Bio similarity The similarity metric is simply the number of common words between the bios of two profiles after removing certain frequently used *stop words* (as is typically done in text retrieval applications). As the set of stop words, we use a popular corpus available for several languages [8].

B. SIMILARITY THRESHOLDS

Clearly the more similar two values of an attribute, the greater the chance that they refer to the same entity, be it a user-name or photo or location. To determine the threshold similarity beyond which two attribute values should be considered as representing the same entity, we rely on *human* annotators. Specifically, we attempt to determine when two attribute values are similar enough for humans to believe they represent the same entity.

We gathered human input by asking Amazon Mechanical Turk (AMT) users to evaluate whether pairs of attribute values represent the same entity or not. We randomly select 200 pairs of profiles and asked AMT users to annotate which attribute values represent the same entity and which do not. We followed the standard guidelines for gathering data from AMT workers [2].

For each attribute, we leverage the AMT experiments to select the similarity thresholds to declare two values as representing the same entity. Specifically, we select similarity thresholds, such that more than 90% of values that represent the same entity (as identified by AMT workers) and less than 10% of the values that represent different entities (as identified by AMT workers) have higher similarities. Consequently, we determine that two user-names or screen-names represent the same name if they have a similarity higher than 0.79, and 0.82 respectively. Two locations represent the same place if they are less than 70km apart. Two photos represent the same image if their SIFT similarity is higher than 0.11 and two bios describe the same user if they have more than 3 words in common.