

The Social Name-Letter Effect on Online Social Networks

Farshad Kooti^{1,*}, Gabriel Magno^{2,*}, and Ingmar Weber³

¹ USC Information Sciences Institute, Marina del Rey, USA

² Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

³ Qatar Computing Research Institute, Doha, Qatar

Abstract. The Name-Letter Effect states that people have a preference for brands, places, and even jobs that start with the same letter as their own first name. So Sam might like Snickers and live in Seattle. We use social network data from Twitter and Google+ to replicate this effect in a new environment. We find limited to no support for the Name-Letter Effect on social networks. We do, however, find a very robust Same-Name Effect where, say, Michaels would be more likely to link to other Michaels than Johns. This effect persists when accounting for gender, nationality, race, and age. The fundamentals behind these effects have implications beyond psychology as understanding how a positive self-image is transferred to other entities is important in domains ranging from studying homophily to personalized advertising and to link formation in social networks.

1 Introduction

According to the *Name-Letter Effect* (NLE), people have a preference for partners, brands, places, and even jobs that share the first letter with their own name. Correspondingly, a Sarah would be more likely to marry a Sam, go to Starbucks, move to San Francisco, and work in sales. This phenomena has been replicated in numerous settings [18,19,9,8,2,20,1] and is part of text books in psychology [13]. Some researchers have, however, questioned the validity or at least the generality of such studies [22,23,14,17,5]. By its supporters, the NLE is usually attributed to “implicit egotism” [20] with people preferring situations that reflect themselves.

We turn to data from online social networks, Twitter and Google+, to see if the NLE can be replicated in a large online setting. Concretely, we seek evidence for or against the NLE in choosing social connections (Sarah following Sam) and in expressing brand interest (Peter following Pepsi). Our findings here are mixed and, depending on the exact setting, we find statistically significant evidence both for and against the NLE.

Extending the NLE and the idea of implicit egotism, we look for a *Same-Name Effect* (SNE) where a Sarah is more likely to follow another Sarah and Tom Cruise is in particular popular among Toms. Here, we observe the presence of the SNE in different settings. We show that the SNE exists for both genders and in different countries. We also show that the SNE affects linking both to celebrities and to normal users and affects

* This work was done while the first two authors were interns at Qatar Computing Research Institute.

both strong and weak ties. Finally, we show that there is an anti-correlation between the number of friends and the extent of link preference bias caused by the SNE.

To the best of our knowledge, this is the first time that the Name-Letter Effect and the Same-Name Effect have been studied in an online setting. It is also the largest study of its kind with more than a million connections analyzed. Our analysis quantifies a factor that affects link formation in online social networks. Understanding the processes governing which links are established is crucial for areas such as information diffusion or link prediction. Moreover, the strength of the NLE or the SNE for an individual could be an estimate of the person's positive self-image. Understanding this could help in understanding homophily, and it could also be used in personalized advertising.

2 Related Work

The NLE was first observed by Nuttin in 1985 [18]. The effect was studied by asking volunteers to pick their favorite letter from pairs or triads of letters where only one of them belonged to the participant's first or last name. Nuttin showed that independent of visual, acoustical, semantic, and frequency characteristics, letters belonging to own first and last name are preferred over other letters. The most popular explanation for the NLE is "implicit egotism" [21]. People have positive feelings about themselves and these feelings are associated implicitly to places, events, and objects related to the self [20].

Later, the presence of the effect was tested in different languages and cultures. It has been shown that the NLE exists in twelve European languages [19]. Also, Hoorens et al., showed that the NLE exists across languages, i.e., participants picked the letters in another language that were either visually or acoustically similar to the letters in their names in their own language [9].

After the discovery of the NLE, many studies verified the existence of the effect in a wide range of decision making situations: People are disproportionately more likely to live in cities and take jobs that are similar to their name [21]. Also, brands that have the same initial as a person's name are preferred by that person [2] and there is a higher chance of donation when the name of the solicitor is similar to the name of the contacted person [1]. Studies have even found that NLE affects marriage; people are more likely to marry a person with a similar name [11,23]. On the other hand, the NLE was not observed in choosing favorite foods and animals [8].

Besides many studies providing evidence for an NLE, there have also been papers questioning the presence of the effect in different areas or the reason for the effect. E.g., in [17], the authors show that a wrong statistical test was conducted in an earlier work on verifying the existence of the NLE in the initial of a baseball player and number of strike outs by him. Also, other works had shown different biases that might create the same results as a NLE [22,23,14]. For example, in the study that showed people are more likely to live in cities with the same initial as theirs, one explanation might be that people in those cities named their babies with such names. Although there are some papers challenging the existence of NLE, the critics are usually concerned about the way a particular study was done, and the main effect is still generally accepted.

3 The Name-Letter Effect on OSNs

In this section, we first test the generality of the NLE on Twitter and Google+ in different domains, such as preferred brands, celebrities, and news media. Then, we investigate the NLE in the social context. Concretely, do users follow other users with the same initial disproportionately more than users with a different initial? Here we use the term “follow” to refer both to Twitter following and to Google+ “has added to a circle”. In both case, the acting user expresses an interest in the updates of the user acted upon.

3.1 Data Description

Twitter: Most of the analyses in this work is done on a large Twitter social network gathered in [3]. The network contains all the 52 million users who joined Twitter by September 2009 and all the 1.9 billion links among them. We also used users’ location information from [12], which uses both location and time zone fields for inferring a user’s country.

Google+: The Google+ dataset was created by collecting public information available in user profiles in the network. The data collection ran from March 23rd of 2012 until June 1st of 2012. In total we were able to retrieve information from 160 million profiles . With the social links of the users, we have constructed a directed graph that has 61 million nodes and 1 billion edges. Details of the Google+ platform and a data characterization of an early version of the dataset is discussed in a previous work [15].

3.2 NLE and Brand Preference

For testing the NLE on Twitter and Google+, we considered a variety of domains and we picked a pair of popular Twitter and Google+ accounts from each domain. Then, we gathered all the followers of each account as of May 2013 (or a large 1 million uniform, random sample of them) in Twitter, and all the followers of each account in Google+ as in the time of the data collection (2012).

We examine the brand NLE by performing the Pearson’s chi-squared test of independence. We do this by counting the followers of each account who have the same initial to see if there are disproportionately many followers for the brands and users with the same initial. For each pair of brands, we create a 2×2 table showing the number of followers for each account whose initial is the same as initial of either of brands. Since both the popularity of the brands and the frequency of name initials are not necessarily the same across the world, in all the analyses in this section we only consider followers in the US. To filter the users in Twitter we used the location field from the users’ accounts and only picked users who had one of the top 20 most populated US cities, *United States*, or *USA* in their location field. The location filter in Google+ was done by extracting the geographic coordinates of the last location present on the *Places lived* field, picking only the users from USA.

Table 1 shows an illustrating example of the 2×2 tables. A represents the number of users who follow Brand 1 and have the same initial as Brand 1. Similarly, D is the number of followers of Brand 2 who have the same initial as Brand 2. For testing the

NLE, first, we calculate the expected values for the cells that the initial of the followers matches the brand’s initial (here A and D). The expected value, is the value that the fields would have if, given the total values, the followers were split uniformly and without any preference. Here the expected value of A would be $\frac{(A+C)*(A+B)}{A+B+C+D}$ and the expected value of D would be $\frac{(B+D)*(C+D)}{A+B+C+D}$. Then, expected values smaller than the observed values for A and D indicate the existence of the NLE.

Table 2 shows all the considered Twitter and Google+ accounts and whether a significant NLE exists or not. We picked these pairs mainly because these accounts have high number of followers. Moreover, the pairs presented here and in the rest of the paper are all the pairs that we did the analysis on, and we are not “cherry-picking” the results. Out of the eight considered domains in Twitter, shown in Table 2, only three of them show a statistically significant NLE, three cases imply NLE but the results are not statistically significant, and the remaining two pairs exhibit a negative NLE. The results suggest that the NLE exists only in some special cases and it is not a generalizable concept for following brands on Twitter. This analysis was done by considering the first name of the user. We repeated the analysis using the Twitter handle (i.e., screen name) of the users. For 61% of the users the initial of the actual name matches the initial of their Twitter handle. Due to this high overlap, testing the NLE by using the handles yields very similar results to using their declared names: in only two cases the results are statistically significant, for the game consoles and the actors, and in these cases the effect is much smaller than the NLE with actual names (3.5% and 1% respectively). In Google+, none of the three pairs of brands/celebrities had statistically significant results, with two of the pairs exhibiting low positive and one negative NLE.

3.3 NLE and Social Link Preference

In this section, we test the NLE in the context of friend link preference. This means that we check if users prefer to establish links to other users with the same initial. To have two sets of users with the same initials for testing the NLE on link preference, we first picked the four most popular names on Twitter that have pairs of same initials: “Michael”, “Matthew”, “Jason”, and “James”. Since these names are used in many countries, considering all users might falsely show the NLE: say “Michael” and “Matthew” are popular in a particular country, but not in others, in this case there will be lots of links from “Matthew” to “Michael”, but not to “Jason”. This could create an

Table 1. Illustration of testing the NLE. If link formation is independent of the initials of the brands, the observed value would be close to the expected value for A , namely, $\frac{(A+C)*(A+B)}{A+B+C+D}$. Larger than expected observed values for A and D indicate the existence of the NLE.

	Brand 1	Brand 2	Total
Brand 1 initial	A	B	$A + B$
Brand 2 initial	C	D	$C + D$
Total	$A + C$	$B + D$	

Table 2. The Twitter and Google+ accounts considered for the brand NLE and the average percentage of preference for the brands with the same initial. There is no significant NLE for most of the brands.

Twitter				Google+			
Account 1	Account 2	NLE	p-value	Account 1	Account 2	NLE	p-value
Sega	Nintendo	9%	< 0.001	Sergey Brin	Larry Page	1%	—
Jim Carrey	Tom Cruise	4%	< 0.001	Nokia	Samsung	-16%	—
Firefox	Internet Explorer	5%	< 0.1	Pepsi	Coca-Cola	1%	—
Canon	Nikon	5%	—				
Puma	Adidas	0.9%	—				
CNN	New York Times	0.4%	—				
Nokia	Samsung	-1.3%	—				
Pepsi	Coca-Cola	-1.7%	—				

Table 3. Results of the NLE on link preference. Effect sizes are shown in the parentheses. In Twitter, users with same initials have negative effect size, contradicting the NLE (p - values < 0.001). Google+ results were not statistically significant.

Twitter				Google+			
	Michael	Jason	Total	Mark	James	Total	
Matthew	6,455 (-2%)	4,285 (+4%)	10,740	Michael	3,605 (0%)	1,829	5,434
James	12,016 (+1%)	7,236 (-2%)	19,252	John	3,213	1,598 (-1%)	4,811
Total	18,471	11,521		Total	6,818	3,427	

apparent NLE in the results, that might not actually exist, or at least not due to implicit egotism. To overcome this issue, we limited ourselves to users in the US.

Table 3 shows the results of the number of times “Matthews” and “Jameses” follow “Michaels” and “Jasons” for Twitter. Surprisingly, the results show a slight, statistically significant¹ *negative* NLE ($\chi^2(1) = 15.58$). This analyses was repeated with a pair of female names (“Melissa” and “Jennifer”) following a pair of male names (“Michael” and “Jason”) and vice versa. Again in both cases a negative NLE existed, but this time not statistically significant. The results clearly show that the NLE does not exist for general social link preference. The same analyses were done for the Google+ dataset, using the two most popular pairs of same initials: “Michael”, “John”, “Marks” and “James”. Again, there was a negative NLE, but not a statistically significant one.

3.4 NLE and Location, Job, and Hobbies

Earlier studies have shown people prefer to live in the cities with the same initials and also choose occupations that have the same initial as their name [21]. We tried to replicate these findings using our data. For Twitter we gathered the profile information of more than 4 million random users and used their location field to see the effect of NLE in the city that people choose to live. For Google+ we retrieved the city from the

¹ In this work, we consider p - value < 0.001 as statistically significant, unless explicitly specified.

“Places lived” field. We tested the NLE for the top ten largest city in the US². The ten largest cities in the US have seven unique initial letters, which leads to 21 (seven choose two), pairs of letters for checking NLE. In Twitter, out of the 21 pairs, 8 pairs show statistically significant results, with 6 of them showing positive NLE. In Google+, 7 pairs were statistically significant, with 6 of them showing positive NLE.

Similarly for the occupations, we consider the following jobs: engineer, cashier, waiter(ess), teacher, and nurse. In Twitter we search the users’ bios for the corresponding strings. The “bio” is the field in the profile that users introduce themselves in and they often include their occupation. In Google+, we examine the “Occupation” field, and looked for the same set of strings. Both in Twitter and Google+ we find only one statistically significant result out of the ten (five choose two) possible pairs of letters, and this single statistically significant pair has negative NLE.

We also test the NLE for the hobbies of the users. More specifically, we look for popular sports in the bio of the users in Twitter and in the “Introduction” field of Google+. We consider football, basketball, baseball, lacrosse, soccer, volleyball, tennis, and hockey. We test the NLE again for the all 21 possible pairs of initials of the sports and the names. Only four of the pairs show a statistically significant result, with only one positive NLE in Twitter and two positive NLE in Google+.

Overall, our findings therefore question the existence or at least the general scope and strength of the NLE as we failed to replicate earlier claims in this new setting.

4 The Same-Name Effect on OSNs

In this section, we test another effect in link creation preference in a more restricted case where both users have the exact same first name, rather than just the same initial. Since all letters of the users’ names are involved, this effect should be stronger than the NLE. We call this effect *same-name effect* (SNE). In other words, are Michaels disproportionately more likely to follow other Michaels compared to other users? A similar idea was tested in an earlier study, where it was shown that people are more likely to marry others with the same *last* name [11]. Here, we analyze linking between users with the same name and show that there is a strong SNE that is robust to many variations.

First, we test the SNE by considering the gender of users as the first name typically identifies the gender. Since men (women) might be more likely to follow other men (women) [16], considering both groups together might cause a false indication of a SNE. So, we perform the SNE test within each gender. Also, as mentioned earlier, having users from different countries might introduce a bias in the results, so again we are considering only users in the US.

We pick the five most popular male names on Twitter among users from the US: “Michael”, “John”, “David”, “Chris”, and “Brian”. Then, we count the number of times each of the users with these names follows other users with these names. Table 4 (appendix) shows the resulting 5×5 table and the effect sizes of 4-13% on Twitter. We calculate the effect size of each name as the average of pairwise preference of that

² http://en.wikipedia.org/wiki/List_of_United_States_cities_by_population

name over other names in the table. This same analysis is repeated in Google+, and the results are the same: male users significantly preferred to follow other users with the same name 7-30% more than expected. We also tested the SNE with the five most popular female names in the US on Twitter and Google+: “Jennifer”, “Jessica”, “Ashley”, “Sarah”, and “Amanda”. The results were similar to the previous case and even stronger: female users significantly preferred to follow others with the same name 30-45% more than expected in Twitter, and 10-29% in Google+.

An alternative explanation for the observed preference could be the fact that different first names are popular in different ethnicities and races. To address this concern, we repeated the analysis for all male first names in the US with more than 10,000 users (56 names in Twitter, 58 names in Google+). We tested the SNE pairwise for these names and the SNE existed for all 1,540 pairs of names with an average effect size of 19% in Twitter, and for all 1,653 pairs of names with an average effect size of 28% in Google+. The fact that the SNE exists for all of the pairs suggests that the preference is not just because of homophily because for at least some cases the names would be associated with the same particular race or ethnicity.

Moreover, we used last names as a proxy of the ethnicity. We used 1990 census data to gather last names that are prominent for only one race in the US.³ We gathered the top 1,000 last names in each of the five races of white, African-American, Asian, Hispanic, and native American natives.⁴ For each race we considered only the last names that are in the top 500 of a particular race and do not occur among the top 1,000 names for any of the other race’s lists. Then we tested the SNE within each race for the pairwise combination of the top 50 popular first names, 1,225 pairs, though not all of these 50 first names were found for all of the five races. Table 5 shows that for all five races a strong and consistent SNE exists in Twitter. In Google+, most of the results were not statistically significant, although implying positive SNE.

To account for age, we use offline data from social security statistics⁵. We focus on the common ages of 20-30 years old on Twitter at the time the data was collected (2009), which corresponds to users born between 1979-1989. We use the records of social security to gather the most popular boy baby names during the mentioned years. Then, we pick all the names that were in the top five at least once: “Michael”, “Jason”, “Christopher”, “Matthew”, “David”, “James”, and “Daniel”. We conduct a similar analysis to the previous section on these names. A statistically significant SNE again existed with 12-17% preference in Twitter, and 5-23% statistically significant preference in Google+. We also try the same experiment with the most popular girl baby names during 1979-1989. Again, a significant SNE is observable with a 16-24% preference in Twitter and 10-106% preference in Google+.

Finally, to see if the SNE exists in different languages and cultures, we picked three countries with different languages: Brazil, Germany, and Egypt. Then, we picked the most popular names in each of those countries and tested the SNE. We found that a statistically significant SNE exists in all three countries, both for Twitter and Google+.

³ <http://names.mongabay.com/>

⁴ Note that in later census the race/ethnicity has been treated differently and that “Hispanic” can now be of any race according to the census terminology.

⁵ <http://www.ssa.gov/oact/babynames/top5names.html>

The effect sizes for Brazilian users range from 13-22% in Twitter and from 16-22% in Google+. Similarly, in Germany and Egypt users significantly preferred to follow other users with the same name (6 - 101%).

5 Discussion

We have focused on testing and observing the NLE and the SNE rather than on explaining them. When using implicit egotism as an explanation the crucial assumption is that users are free to choose the brands they like or the members of their social network. This basic assumption is arguably flawed as people can only connect to people (or brands) they know. But as the distribution of names is not homogeneous across all parts of society this creates implicit selection biases. For example the name Emma was very popular for girls born during 2002-2012⁶ but less popular earlier which, in turn, means that an Emma would be more likely to go to school with another Emma and hence have a chance to connect. Similarly, the name DeShawn is popular among African Americans [6] which means a DeShawn growing up in a predominantly black neighborhood would again have a higher than expected chance of connecting to another DeShawn. In fact, previous research has shown that mere familiarity with a name correlates with likeability [4,7].

We tried to avoid obvious pitfalls, such as selecting names associated with a particular demographic groups, and we looked at names that were popular during a certain period. Additionally, the fact that for testing the NLE and the SNE on link preference we only used the network of early adopters of Twitter (up to September 2009) and Google+ (less than a year after the launch) helps to further homogenize the user set across age and income. Also, we have used users' last names to test the SNE within one race. Still, naming conventions within a family, where family members are given the same first name, could explain part of the observed the SNE.

It is also not clear what fraction of users use their real name in online social networks. We believe this is the case for the majority of the users, especially for Google+, since Google explicitly asked users to use their real name and banned the accounts of users with fake names⁷. There might be much less use of real names on Twitter, but the fact that our findings for Twitter and Google+ are very consistent suggests that there is no dramatic difference between Twitter and Google+ in the way people chose their name. And even if the majority of the names are not real, we still found the SNE, which might have a different explanation than the implicit egotism. Also, note that for testing SNE, we tested the effect on common English names, so we are not analyzing users completely fake names like "cowboy".

6 Conclusions

The Name-Letter Effect (NLE) states that people prefer the letters in their own names over other letters. We investigated the existence of the NLE in the context of Twitter

⁶ <http://www.ssa.gov/oact/babynames/top5names.html>

⁷ <http://gawker.com/5824622/names-banned-by-google-plus>

and Google+. Our findings question at least the generality of the NLE. Going beyond the NLE, we analyzed users' linking behavior for a same-name effect (SNE), where instead of comparing the initials we compared the whole name. In this stronger version, we observe a robust effect, even when accounting for gender, age, race, and location.

Besides the psychological aspects of NLE, there are some real-world implications. E.g., one study has showed that using NLE can increase the chance of donation made by people [1]. In recent years, the Coca-Cola *share a coke*⁸ campaign has proven to be very successful by increasing sales⁹.

References

1. Bekkers, R.: George gives to geology jane: The name letter effect and incidental similarity cues in fundraising. *International Journal of Nonprofit and Voluntary Sector Marketing* 15(2), 172–180 (2010)
2. Brendl, C.M., Chattopadhyay, A., Pelham, B.W., Carvallo, M.: Name letter branding: Valence transfers when product specific needs are active. *Journal of Consumer Research* 32(3), 405–415 (2005)
3. Cha, M., Haddadi, H., Benevenuto, F., Gummadi, K.: Measuring User Influence in Twitter: The Million Follower Fallacy. In: ICWSM (2010)
4. Colman, A.M., Hargreaves, D.J., Sluckin, W.: Psychological factors affecting preferences for first names. *Names: A Journal of Onomastics* 28(2), 113–129 (1980)
5. Dyjas, O., Grasman, R.P., Wetzels, R., van der Maas, H.L., Wagenmakers, E.J.: What's in a name: a bayesian hierarchical analysis of the name-letter effect. *Frontiers in Psychology* 3 (2012)
6. Fryer, R.G., Levitt, S.D.: The causes and consequences of distinctively black names. *The Quarterly Journal of Economics* 119(3) (2004)
7. Hargreaves, D.J., Colman, A.M., Sluckin, W.: The attractiveness of names. *Human Relations* 36(4) (1983)
8. Hodson, G., Olson, J.M.: Testing the generality of the name letter effect: Name initials and everyday attitudes. *Personality and Social Psychology Bulletin* 31(8), 1099–1111 (2005)
9. Hoorens, V., Nuttin, J.M., Herman, I.E., Pavakanun, U.: Mastery pleasure versus mere ownership: A quasi-experimental cross-cultural and cross-alphabetical test of the name letter effect. *European Journal of Social Psychology* 20(3) (1990)
10. Johnson, B., Eagly, A.: Effects of involvement on persuasion: A meta-analysis. *Psychological Bulletin* 106, 290–314 (1989)
11. Jones, J.T., Pelham, B.W., Carvallo, M., Mirenberg, M.C., et al.: How do i love thee? let me count the js: Implicit egotism and interpersonal attraction. *Journal of Personality and Social Psychology* 87, 665–683 (2004)
12. Kulshrestha, J., Kooti, F., Nikravesh, A., Gummadi, P.K.: Geographic dissection of the twitter network. In: ICWSM (2012)
13. Leary, M.R., Tangney, J.P. (eds.): *Handbook of Self and Identity*, 2nd edn. The Guilford Press (2011)
14. Lebel, E.P., Paunonen, S.V.: Sexy but often unreliable: The impact of unreliability on the replicability of experimental findings with implicit measures. *Personality and Social Psychology Bulletin* 37(4), 570–583 (2011)

⁸ <http://www.coca-cola.co.uk/faq/products/share-a-coke.html>

⁹ 3MM P4W Consumption Oct-Dec 2011 B3 Survey Australia

15. Magno, G., Comarela, G., Saez-Trumper, D., Cha, M., Almeida, V.: New kid on the block: exploring the google+ social graph. In: Proceedings of the 2012 ACM Conference on Internet Measurement Conference, pp. 159–170. ACM, New York (2012)
16. Magno, G., Weber, I.: International gender differences and gaps in online social networks. In: SocInfo (2014)
17. McCullough, B., McWilliams, T.P.: Baseball players with the initial “k” do not strike out more often. *Journal of Applied Statistics* 37(6), 881–891 (2010)
18. Nuttin, J.M.: Narcissism beyond gestalt and awareness: The name letter effect. *European Journal of Social Psychology* 15(3), 353–361 (1985)
19. Nuttin, J.M.: Affective consequences of mere ownership: The name letter effect in twelve european languages. *European Journal of Social Psychology* 17(4), 381–402 (1987)
20. Pelham, B.W., Carvallo, M., Jones, J.T.: Implicit egotism. *Current Directions in Psychological Science* 14(2), 106–110 (2005)
21. Pelham, B.W., Mirenberg, M.C., Jones, J.T.: Why susie sells seashells by the seashore: implicit egotism and major life decisions. *J. Pers. Soc. Psychol.* 82(4), 469–487 (2002), <http://www.biomedsearch.com/nih/Why-Susie-sells-seashells-by/11999918.html>
22. Simonsohn, U.: In defense of diligence: a rejoinder to pelham and carvallo. *Journal of Personality and Social Psychology* 101(1), 31–33 (2011)
23. Simonsohn, U.: Spurious? name similarity effects (implicit egotism) in marriage, job, and moving decisions. *Journal of Personality and Social Psychology* 101(1), 1–24 (2011)

A Appendix

Table 4. Results for the SNE on popular names in the US on Twitter and Google+. The effect sizes are positive for all five names showing the SNE ($p - values < 0.001$).

	Twitter					
	Michael	John	David	Chris	Brian	Total
Michael	28,587 (+5%)	36,590	29,051	25,928	15,093	135,249
John	28,393	42,417 (+4%)	31,540	27,823	16,906	147,079
David	24,303	33,713	29,388 (+5%)	24,441	14,255	126,100
Chris	22,632	31,383	25,107	25,999 (+6%)	14,089	119,210
Brian	15,394	20,974	16,676	15,636	11,383 (+13%)	80,063
Total	119,309	165,077	131,762	119,827	71,726	

	Google+					
	Michael	David	John	Chris	James	Total
Michael	5,949 (+7%)	4,492	4,659	5,349	1,829	22,278
David	4,739	5,375 (+10%)	4,526	4,858	1,657	21,155
John	4,590	3,979	5,154 (+9%)	4,687	1,598	20,008
Chris	3,971	3,431	3,659	4,791 (+9%)	1,410	17,262
James	2,349	1,980	2,329	2,556	1,287 (+30%)	10,501
Total	21,598	19,257	20,327	22,241	7,781	

Table 5. The SNE test for users with a race-specific last name. The “# of last names” indicates the number of race-specific last names found. “Could be tested” is the number of first name pairs where each first name had a non-zero count for the race-specific last name. There is a large number of statistically significant positive effects, and only a single first name pair with a significant *negative* effect.

Twitter							
Race	# of last names	Could be tested	Median SNE	Sig. positive	Sig. negative	Non-sig. positive	Non-sig. negative
White	35	1225	45%	986	0	237	2
Asian	394	1193	59%	433	1	645	114
Hispanic	341	1073	86%	350	0	541	146
Native American	72	345	100%	78	0	221	46
African-American	64	263	100%	56	0	241	66

Google+							
Race	# of last names	Could be tested	Median SNE	Sig. positive	Sig. negative	Non-sig. positive	Non-sig. negative
White	35	80	43%	0	0	79	1
Asian	394	608	47%	61	0	494	53
Hispanic	341	95	60%	0	0	95	0
Native American	72	1	-17%	0	0	0	1
African-American	64	16	18%	0	0	16	0

A.1 SNE and Social Tie Strength

We also investigated the correlation of the SNE and the strength of the tie between users. Concretely, are users’ strong ties more affected by the SNE than their weak ties? Again, we limited this analysis to users from the US and the mentioned popular names of American users on Twitter.

We eliminated all the super-users to better capture the strong and weak ties among normal users. For all normal users, for the link from user A to B , we looked at the Jaccard similarity of the friends of A and B as a measure of the strength of the tie. Then, we considered half of the links with the lower strength as weak links and the other half as strong ties ($threshold = 0.008$). First, we tested the SNE by only considering weak ties, and then by only considering strong ties. In both cases, the SNE was statistically significant. For weak links the preference ranged from 13% to 17% and for strong ties from 10% to 13%, and for all the five names the SNE was slightly stronger for the weak ties. Our results suggest that people are more affected by SNE when they are establishing a weak link. This is in contrast with an earlier study that has found the NLE only affects people’s important decisions, such as choosing a job or place to live, and not the more trivial decisions like favorite animals or foods [8]. This observation was explained by an earlier finding that the NLE is a type of implicit egotism and implicit egotism is boosted under stress [10]. However, we do not find evidence to support this finding in Twitter. Though, the results on Google+ are not consistent with these findings and the preference for weak ties ranged from 7-33% and for strong ties

12-45%. Further investigation of differences between Twitter and Google+ is needed to figure out the root of the mentioned inconsistency.

A.2 SNE and Number of Friends

Finally, we examined the correlation between the SNE and the number of friends (followers) of users. The aim is to see if the SNE differs for users with more compared to users with less friends. Similar to before, we considered only users from the US and the mentioned popular names on Twitter. Then, we grouped users based on their number of friends logarithmically, up to 64 friends and a group for users with more than 64 friends. The resulting groups are fairly balanced, with the smallest group (one friend) containing 8% of considered users and the largest group (between 16 and 32 friends) 20% of them. We also use the same group sizes for Google+.

We tested the SNE in each of the groups by only considering the links going out from users of that group and then taking the average of the SNE for the five considered names. Figure 1 shows that there is a noticeable reverse correlation between the number of friends and the SNE. Users with fewer friends are more likely to follow other users with the same name compared to the users with a higher number of friends.

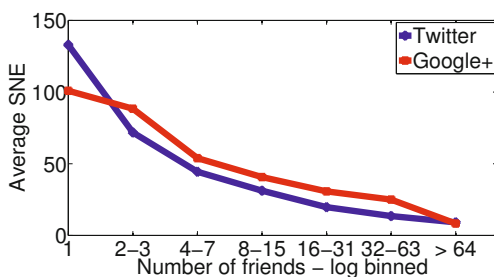


Fig. 1. The average SNE of users grouped by the number of friends