All questions should be answered in your own words. You may paraphrase the authors and include no more than two direct quotes from the source papers. Your write-up should be 1-2 pages long, with reasonable (e.g. 12 pt) font and margins.

1. What is the key need for this algorithm/approach/system?

2. What benefits does this algorithm/approach/system provide over existing algorithms/approaches/systems?

3. What is the key object? (e.g. tweet, RDD, Tensor)

4. Describe a technical challenge in implementing the author's solution

5. What is a key insight or lesson learned from the paper?

6. What is something interesting your learned from this paper, or your thoughts about its strengths and/or weaknesses. Is there anything else interesting about this paper, or your interpretation of it that you want to share?

7. Provide a citation for the additional paper you read. How does this paper relate to the assigned one? What new information did you learn by reading it?

Grading Rubric (10 points total)
1. (5) Were the questions answered, based on the papers' content, in the student's own words?
2. (3) Is an additional paper discussed, and is it relevant?
3. (2) Was the paper well-written - proper grammar, no misspellings, etc.?

Example

Mahmud J, Nichols J, Drews C. Home location identification of twitter users. ACM Transactions on Intelligent Systems and Technology (TIST). 2014;5(3):47.

1. Location information can help determine the location of described events and users, how widespread an event is, geographically, and can help determine who might be interested in hearing about the information. However, less than 1% of tweets contain geographic location information and the user's profile location is often not filled in or may no longer be accurate or applicable.

2. This approach leverages multiple sources to determine a user's home location, at the city level. These sources include explicit location information as well as information derived from the tweet's content and the user's tweet history. In particular, they use explicit mentions of locations in the tweet text, time zone information, and ensemble methods to achieve high accuracy.

3. The fundamental concepts here are tweets and users. The goal of this research is to determine a user's home location based on tweet content, history and metadata. Tweets are short messages broadcast on Twitter either directly or via a third party application, such as foursquare.

4. A key challenge in this work was finding algorithm(s) that worked on the varying amount of data available for each user. The authors experimented with different ensemble methods to find one (dynamically weighted) that performed well.

5. This paper brings to light the possibility of determining a person's location without their explicit permission. In addition, the truth was determined by collecting geo-tagged tweets. It's possible that twitter users who geo-tag their tweets have different location & tweeting patterns than those who do not geo-tag their tweets. In addition, tweeters with private profiles were excluded from consideration. Another question is what about business tweeters? It's possible that more than one user might be using the same account.

6. Accuracy is, in many ways, a controversial evaluation metric. It is often easy to achieve high accuracy merely by guessing the most common value. It also doesn't take in to account how far off an incorrect answer is. For example, guessing Austin, Texas instead of Houston, Texas is, presumably, a more accurate result than predicting New York City, New York in lieu of Houston. The authors used a soft distance measure, but it works in their favor - counting locations as correct if they are within a certain distance of the true location. Being widely off doesn't hurt them any more than being slightly off-target.

7. Compton R, Jurgens D, Allen D, editors. Geotagging one hundred million twitter accounts with total variation minimization. Big Data (Big Data), 2014 IEEE International Conference on; 2014: IEEE.

This paper takes a network optimization approach to determining user locations, instead of a content based approach. The authors build a network of tweeters (nodes) and reciprocated mentions (edge weights). They then iteratively infer the unknown locations, until convergence. The advantage of this approach is that it avoids language issues and "noisy" tweets. It also scales well. However, a key limitation of this method is that it can only by evaluated with users whose locations are known. Other studies (Venturini 2017) show that users who geocode their tweets are not representative of tweeters overall. In addition, this approach is limited to city level granularity.