Pittsburgh Emoji and Word Map

By Jennifer Chou and Dan Tasse

**What do people in Squirrel Hill talk about?**

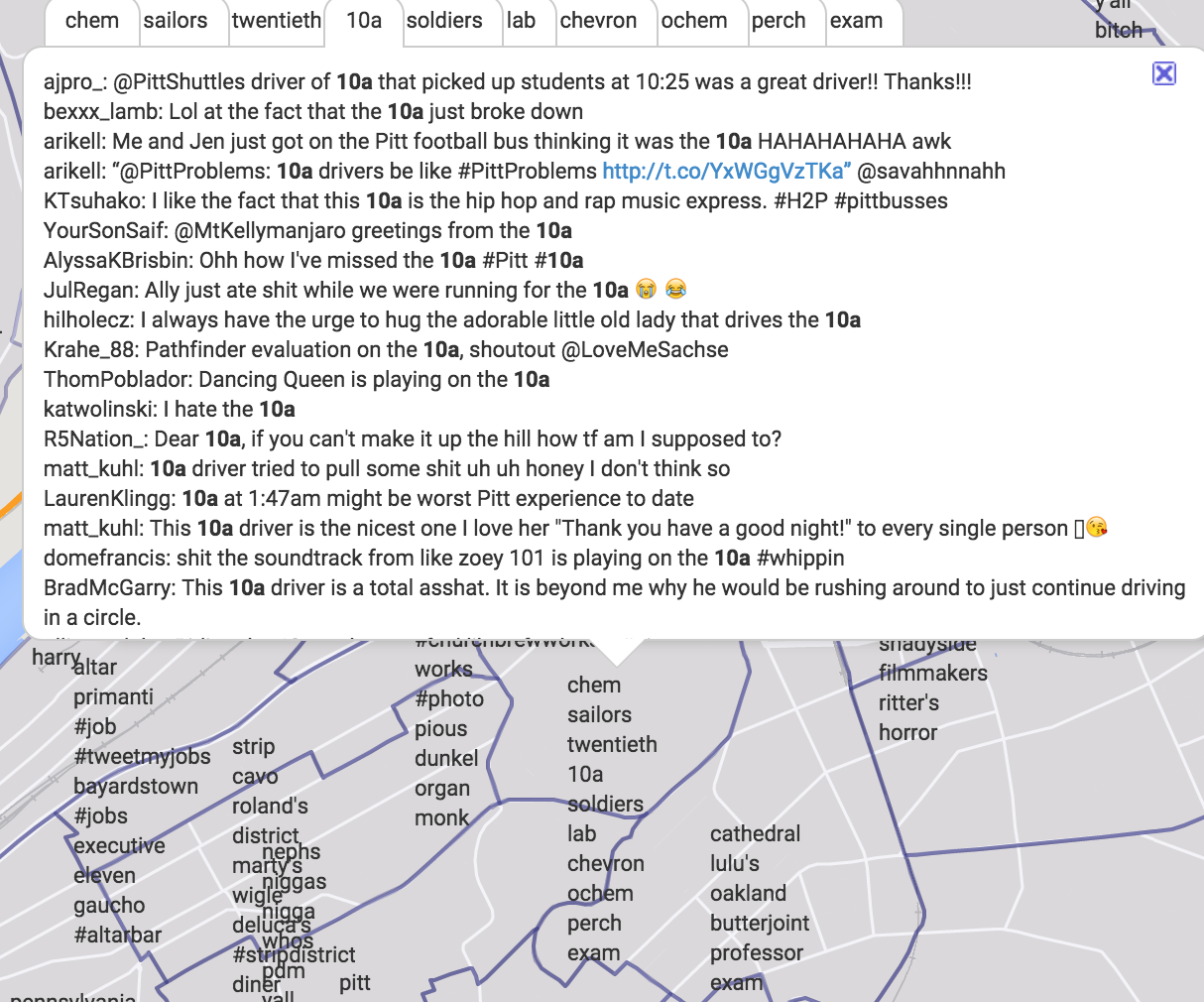
Or, more interestingly, what do people in Squirrel Hill talk about that people in other neighborhoods don’t? What is it that makes Squirrel Hill Squirrel Hill? That’s the question we set out to answer with this project.



**How it works**

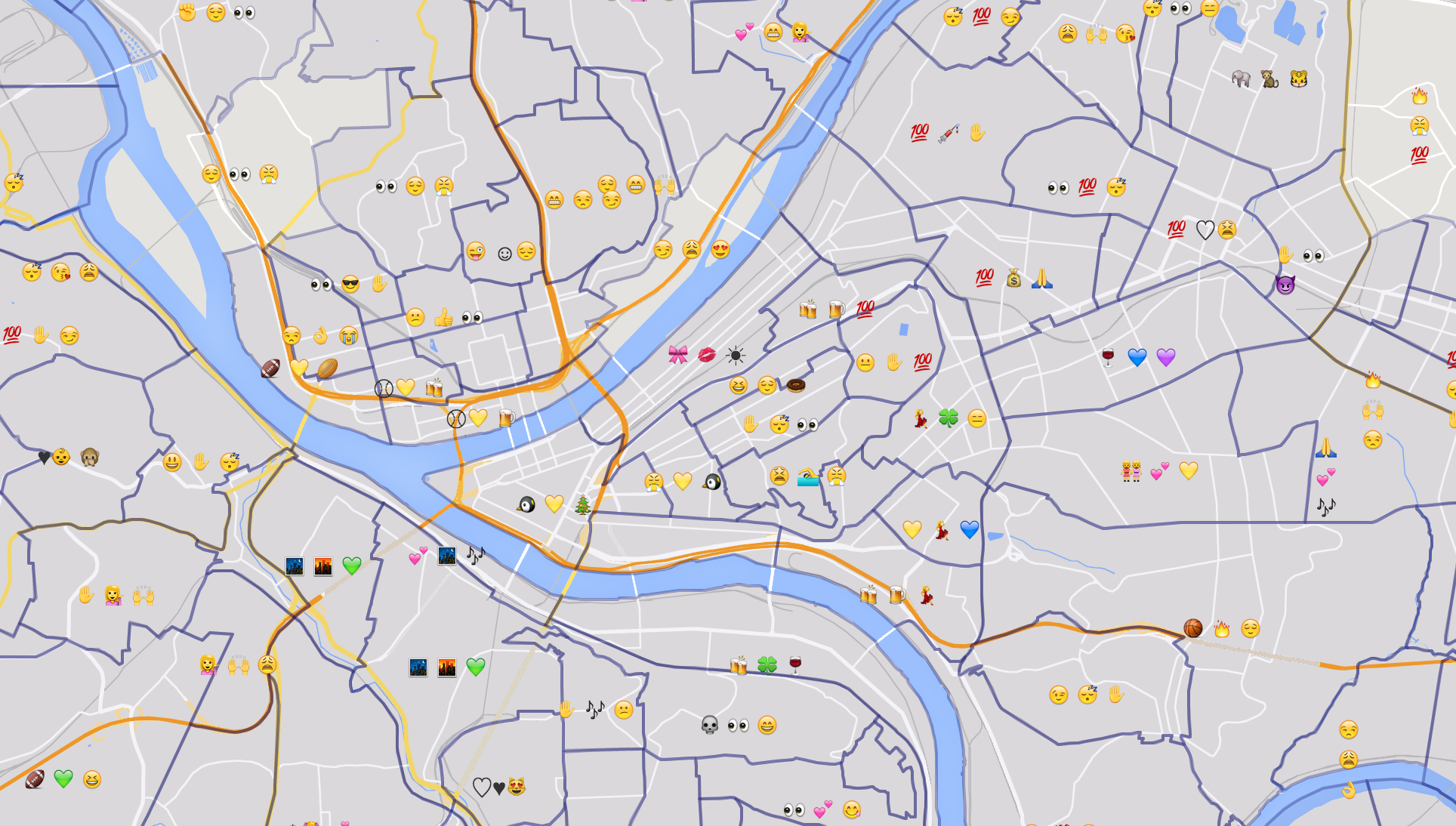
We gathered all tweets geotagged in Pittsburgh over about a year, from December 2013 to January 2015. We sorted them by neighborhood (using boundaries provided by the WPRDC) and used a modified TF-IDF algorithm to figure out what words were specific to each neighborhood. This algorithm counts the frequency of a word in a given neighborhood, and then adjusts the word’s final score based on how many other neighborhoods also use that word.

For example, “Steelers” is used a lot in Squirrel Hill, but it’s also used in many other neighborhoods, so it has a pretty low score. “Tunnel”, however, is quite popular in Squirrel Hill (mostly due to people grousing about tunnel traffic), but not elsewhere. Similarly, “10a” is a popular bus used to get around Pitt, but isn’t used elsewhere, so “10a” shows up a lot in Oakland.



**Emoji**

These words just represent what people are talking about on Twitter. What are people feeling? For that question, we looked to the emojis people are tweeting. Emojis are an interesting new form of communication: often one character can say more than a word, so they can tell us about where people like to do certain things, or maybe even how people feel.



For example, we can see that the zoo is up in Highland Park, and that people like watching baseball and football and drinking beer on the North Shore. Obvious enough. But did you know how popular the swimming pool in Oakland is, or the Christmas tree lighting downtown?

**Future work, and so what?**

There’s still work to do, of course. One major challenge is algorithmic: How do we combine these posts from multiple people into a representative aggregate? A lot of these words/emojis are boosted by one person tweeting them multiple times. We don’t want one person to dominate the neighborhood’s tweets, but we do want an avid basketball fan to count more than someone who just tweeted about basketball once.

We hope this is the first step towards useful neighborhood guides. Imagine if you were moving to Pittsburgh for the first time, and looking for the right neighborhood to live in. Knowing that Squirrel Hill South has a lot of basketball fans, or that the top words in Lawrenceville are all trendy bars or music venues, could really help you get a feel for the city and its parts.

Try it out! http://emojimap.herokuapp.com (be patient; it’s on a free server so it’ll be a little slow.) And send any feedback or ideas to dantasse@cmu.edu.