Density and Hubness

Libby Beer (libby.beer@gmail.com)
Jesse Metcalf-Burton (datasciencejess@gmail.com)
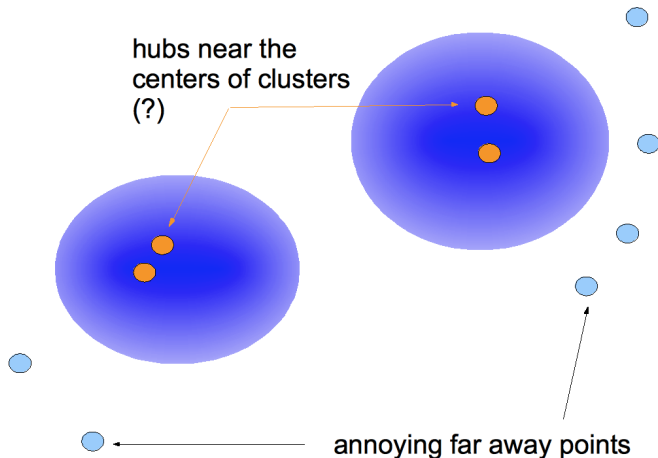Marilyn Vazquez ()

July 20, 2017

Libby goes here

Can hubs help us cluster?

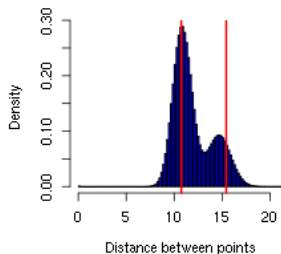

hubs near the
centers of clusters
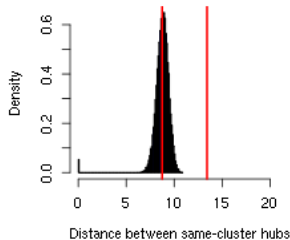(?)

annoying far away points

Distances with synthetic data

n-hubness//users/jesse//graphs/gaussians_1000_5000_60.mat.10.png
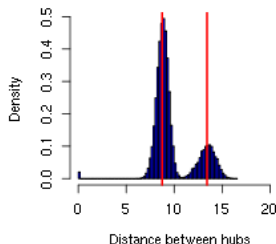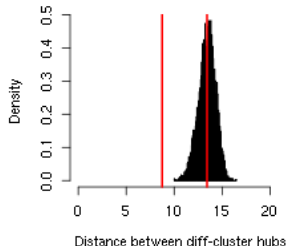
Distances with (other) synthetic data

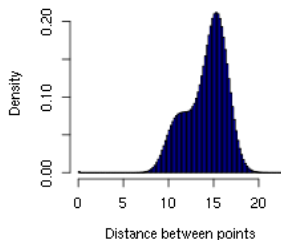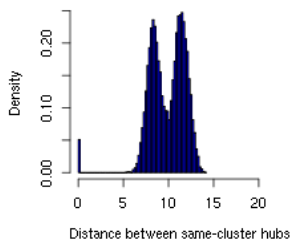/ubness//users/jesse//graphs/synthetic_data_30_50_overlap30.mat.10.png
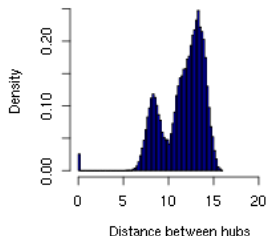
Distances with real data

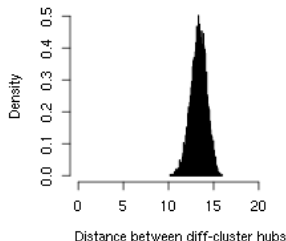:7/wisdm-hubness//users/jesse//graphs/data_spam.mat.10.png
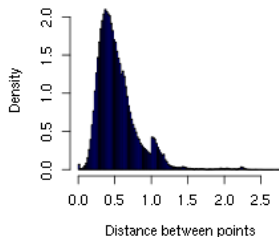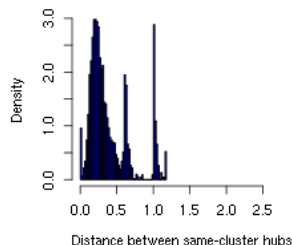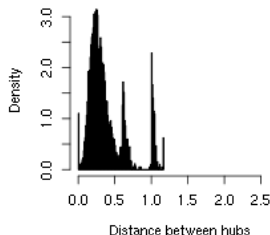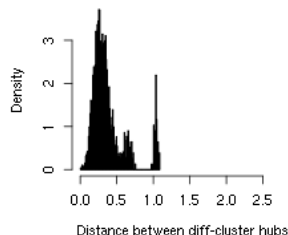
Do hubs appear in high density regions?

Intuition

Do hubs appear in high density regions?

Answer: Sometimes!

gauss, dim=60, dens=1to2, k=10
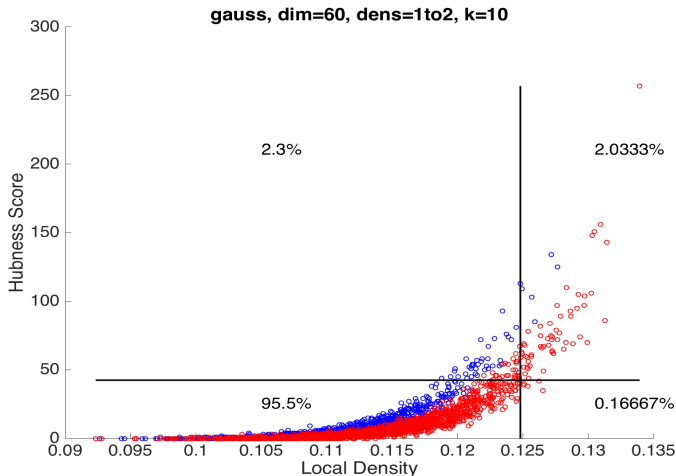
Notice: (1) Proportion of hubs in the two density regions, (2) proportion of high density points that are hubs, and (3) more hubs in the denser cluster.

Do hubs appear in high density regions?

Answer: Sometimes!

Notice: (1) A LOT more hubs in the denser cluster and (2) the rise of two "wings" that correspond to the two different cluster.

Observations

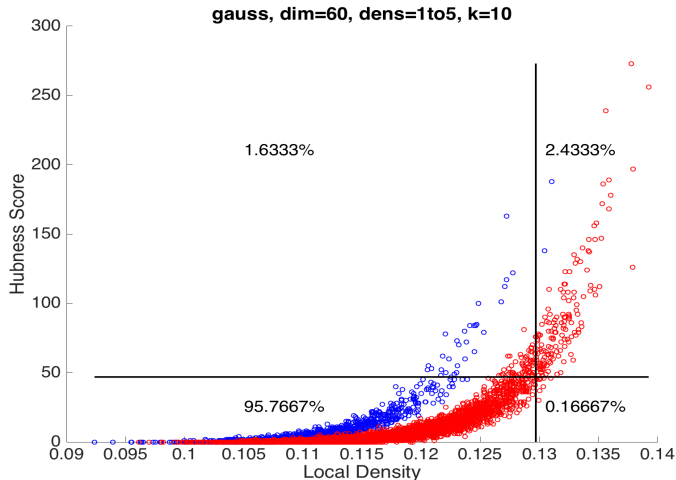From the experiments on the Gaussian data, we observe that:

- About 13 out of 36 experiments, there was a larger proportion of hubs in the low density region i.e. hubness scores and density are not exactly the same thing
- About 23 out of 36 experiments, there was a larger proportion of hubs in the high density region i.e. if a point is a hub, it is more likely to be in a high density region.
- All our experiments show that high density regions are mostly composed of hubs i.e. if a point is in a high density region, it is most likely a hub.

Future Work

- More experiments of local density versus hubness scores for other density types e.g. uniform density
- Verify if patterns also show up in real data
- Data with more clusters to see if "wings" arise for each cluster