

Density and Hubness

Libby Beer (libby.beer@gmail.com)

Jesse Metcalf-Burton (datasciencejess@gmail.com)

Marilyn Vazquez ()

July 20, 2017

Libby goes here

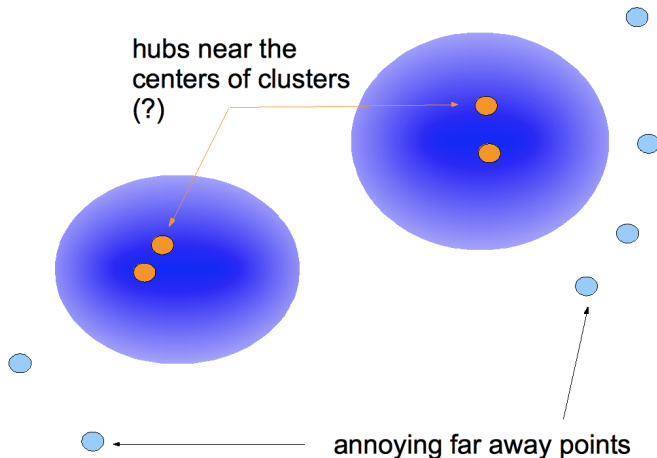
Project 5: The Hubness Phenomenon in High Dimensional Spaces

- Carlotta Domeniconi (George Mason University)
- Sibel Tari (Middle East Technical University)
- Gulce Bal (Middle East Technical University)
- Libby Beer (Institute for Defense Analyses)
- Hillary Fairbanks (University of Colorado Boulder)
- Priya Mani (George Mason University)
- Jesse Metcalf-Burton (Department of Defense)
- Marilyn Vazquez (George Mason University)

Group 5 Overview

Distances between points? Hubs?

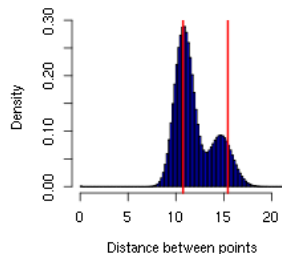
Can hubs help us cluster?



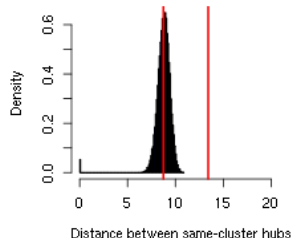
Distances with synthetic data

1-hubness//users/jesse//graphs/gaussians_1000_5000_60.mat.10.png

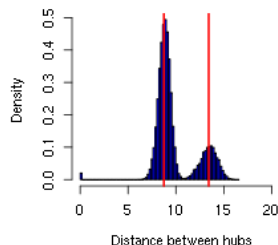
Distances between (all) points



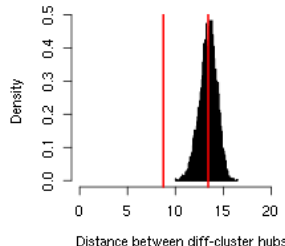
Intra-cluster distances



Distances between hubs



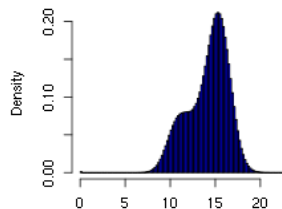
Inter-cluster distances



Distances with (other) synthetic data

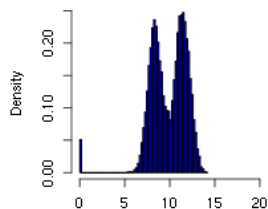
libness//users/jesse//graphs/synthetic_data_30_50_overlap30.mat.10.png

Distances between (all) points



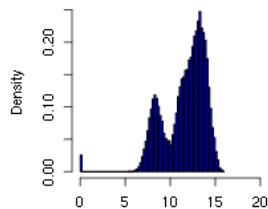
Distance between points

Intra-cluster distances



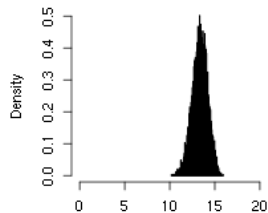
Distance between same-cluster hubs

Distances between hubs



Distance between hubs

Inter-cluster distances

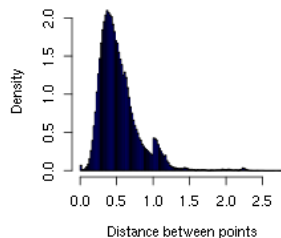


Distance between diff-cluster hubs

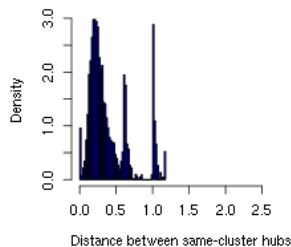
Distances with real data

:/wisdm-hubness//users/jesse//graphs/data_spam.mat.10.png

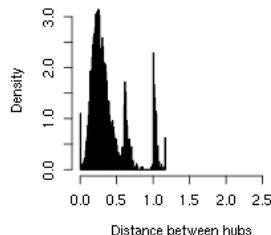
Distances between (all) points



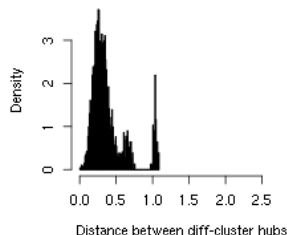
Intra-cluster distances



Distances between hubs

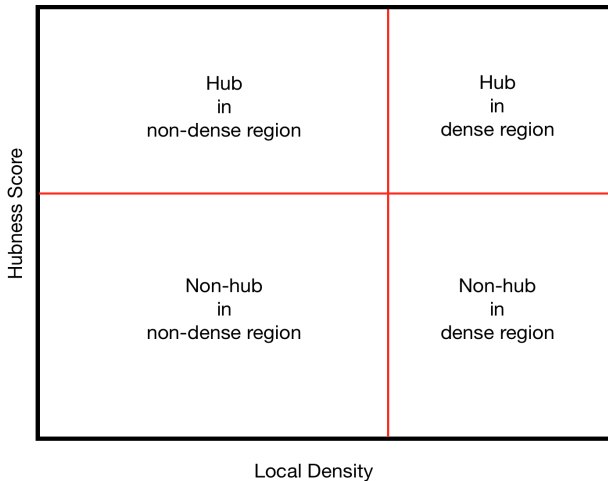


Inter-cluster distances



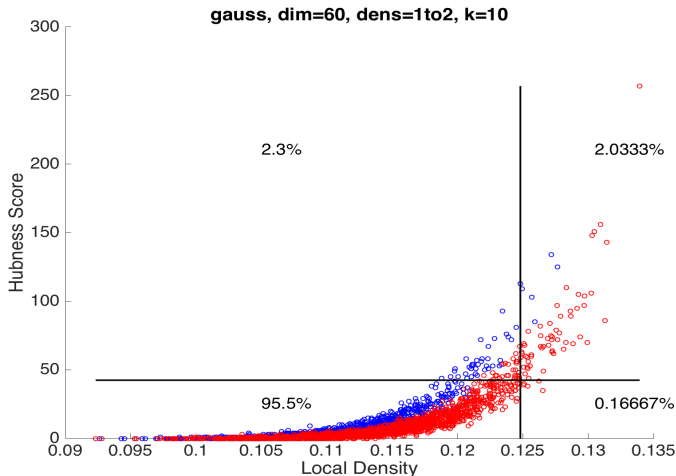
Do hubs appear in high density regions?

Intuition



Do hubs appear in high density regions?

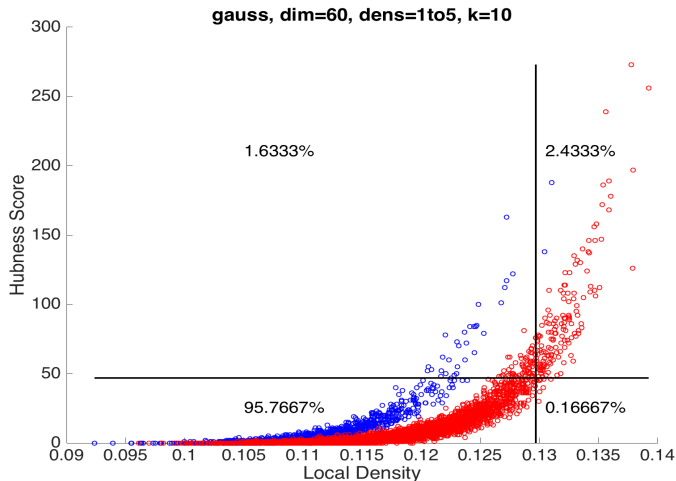
Answer: Sometimes!



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.

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.

- 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