

Visual Analysis of Thematic, Social and Geospatial Patterns of Microblogging Content Using D3

Thomas Gründemann, Dirk Burghardt

AGILE 2016, Helsinki, 14 June 2016



**DRESDEN
concept**
Exzellenz aus
Wissenschaft
und Kultur

Content

1 Introduction

2 Data Sets and Tools

3 Analysis and Visualization of **Thematic Relations**

4 Analysis and Visualization of **Social Relations**

5 Analysis and Visualization of **Spatial Distributions**

1 Introduction

- Social network platforms are a popular media for the communication over the internet
- Conversation creates patterns with identifiable structures
- **User Generated Content** often have direct or indirect
 - thematic reference
 - social reference
 - geospatial reference
 - temporal reference
- Visual analysis tools can give an insight into the structure of the data



2 Data Set and Tools

Data Set 1: Soccer	Data Set 2: Politics
<ul style="list-style-type: none">• search query with the hashtag #fcbsvw on 20 April 2016• #fcbsvw was the keyword for the soccer match between FC Bayern Munich and SV Werder Bremen on 19 April 2016• Contains 624 tweets• Time period 19 - 20 April 2016	<ul style="list-style-type: none">• Data stream of tweets with the hashtag #pegida• Pegida is a xeno- and islamophobic protest movement founded in Dresden, Germany and is demonstrating weekly on Monday in this city• Contains 381 tweets• Time period of 24 hours around the Monday demonstration from 18 January 2016

2 Data Set and Tools

Tools for analyzing and visualizing the data set

- Data analysis with Python
- Data visualization with D3
- D3
 - JavaScript library for producing dynamic and interactive data visualization
 - Uses SVG, HTML5 and CSS standards

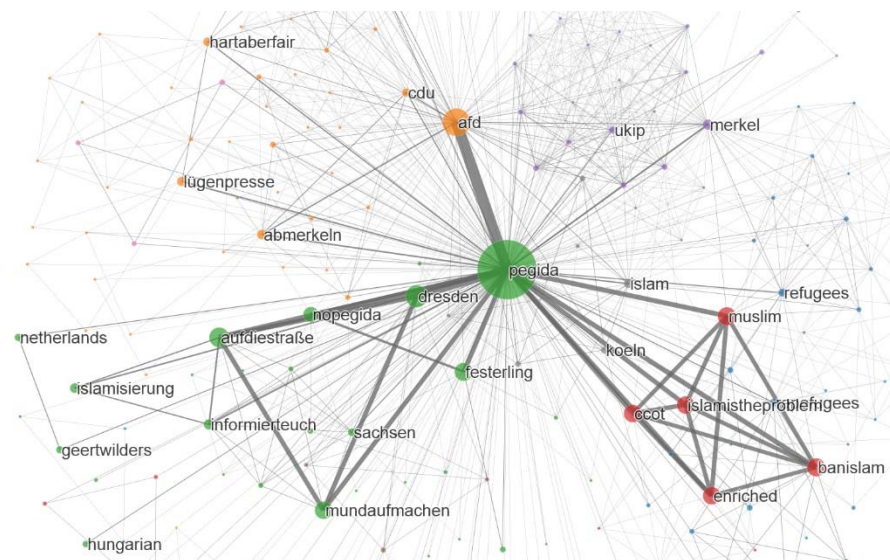
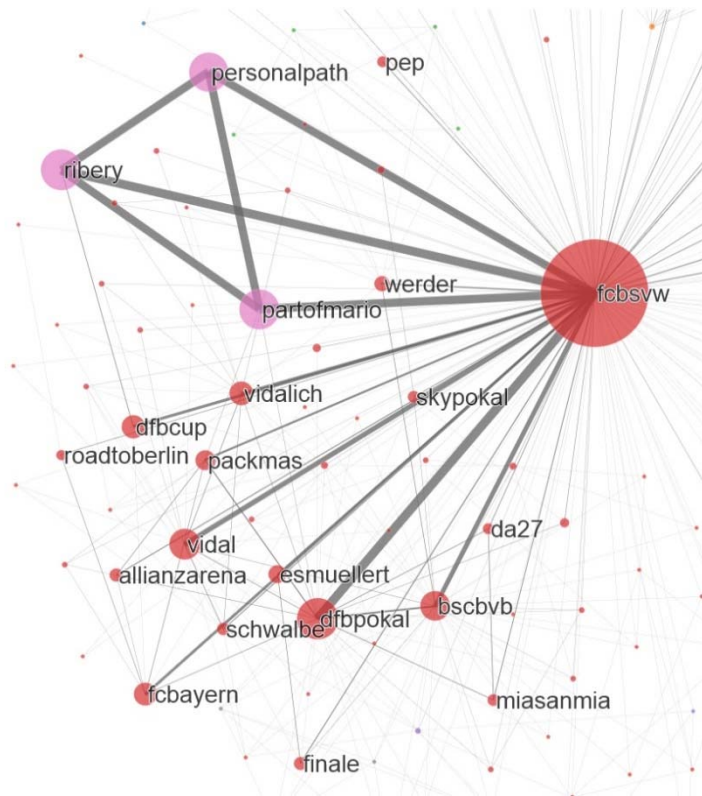


3 Analysis and Visualization of Thematic Relations

- By adding hashtags to the message, users build a semantic network linking their messages
- The network can be represented as a graph:
 - nodes represent hashtags
 - edges represent the co-occurrence of hashtags within a single tweet
- Examine the hashtags that occur together in a tweet, can be a first step to discover topics

3 Analysis and Visualization of Thematic Relations

Soccer #fcbsvw | Politics #pegida



- 156 different hashtags (nodes)
- 509 different hashtag pairs (edges)
- 218 different hashtags (nodes)
- 905 different hashtag pairs (edges)

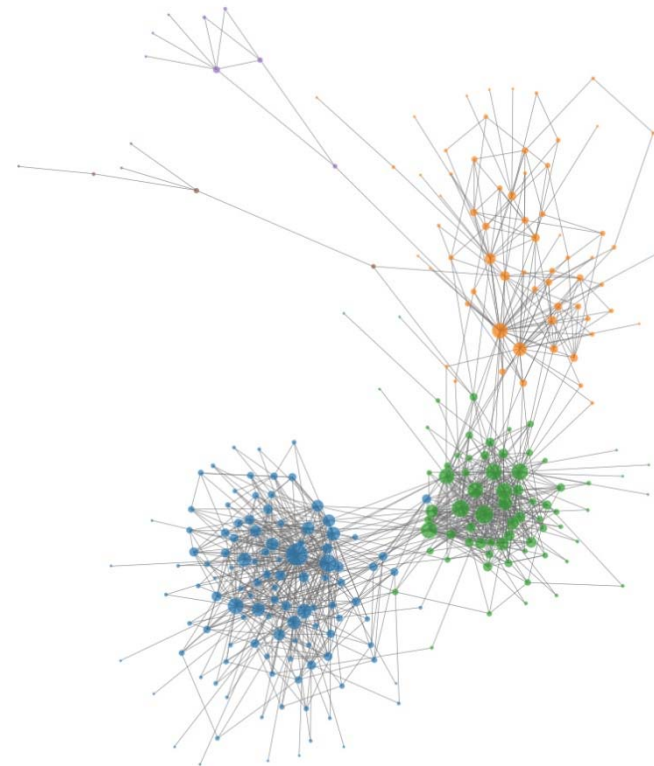
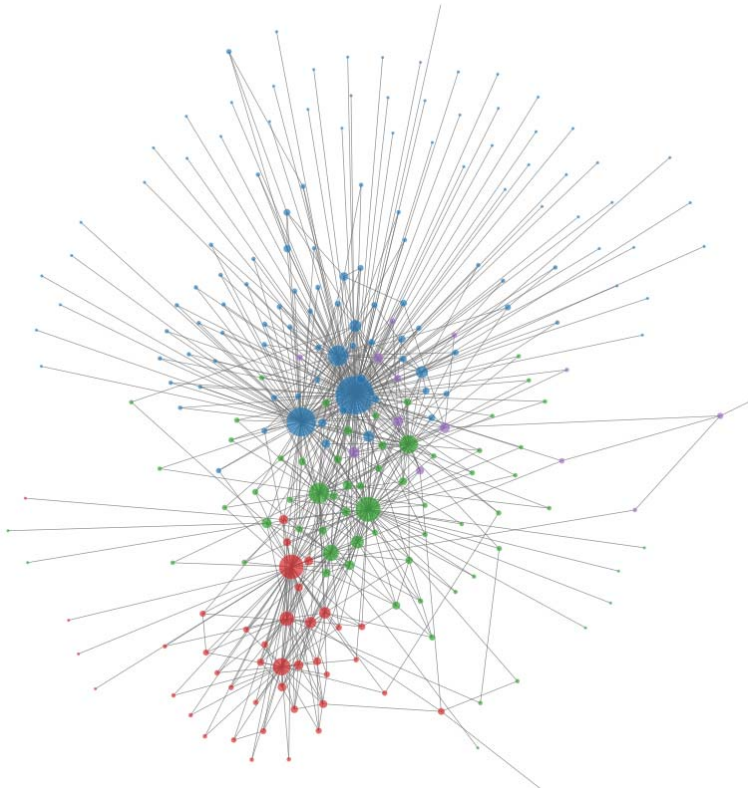
4 Analysis and Visualization of Social Relations

- Users are connected by conversation
- On Twitter, this is realized as follower or friend
 - Followers are users who follow “me”
 - Friends are users who “I” follow
- The visualization of the user relationships in a network gives an insight into the structure of the conversation

4 Analysis and Visualization of Social Relations

Soccer #fcbsvw

Politics #pegida



- 358 different users (nodes)
- 739 links between users (edges)

- 298 different users (nodes)
- 1,080 links between users (edges)

4 Analysis and Visualization of Social Relations

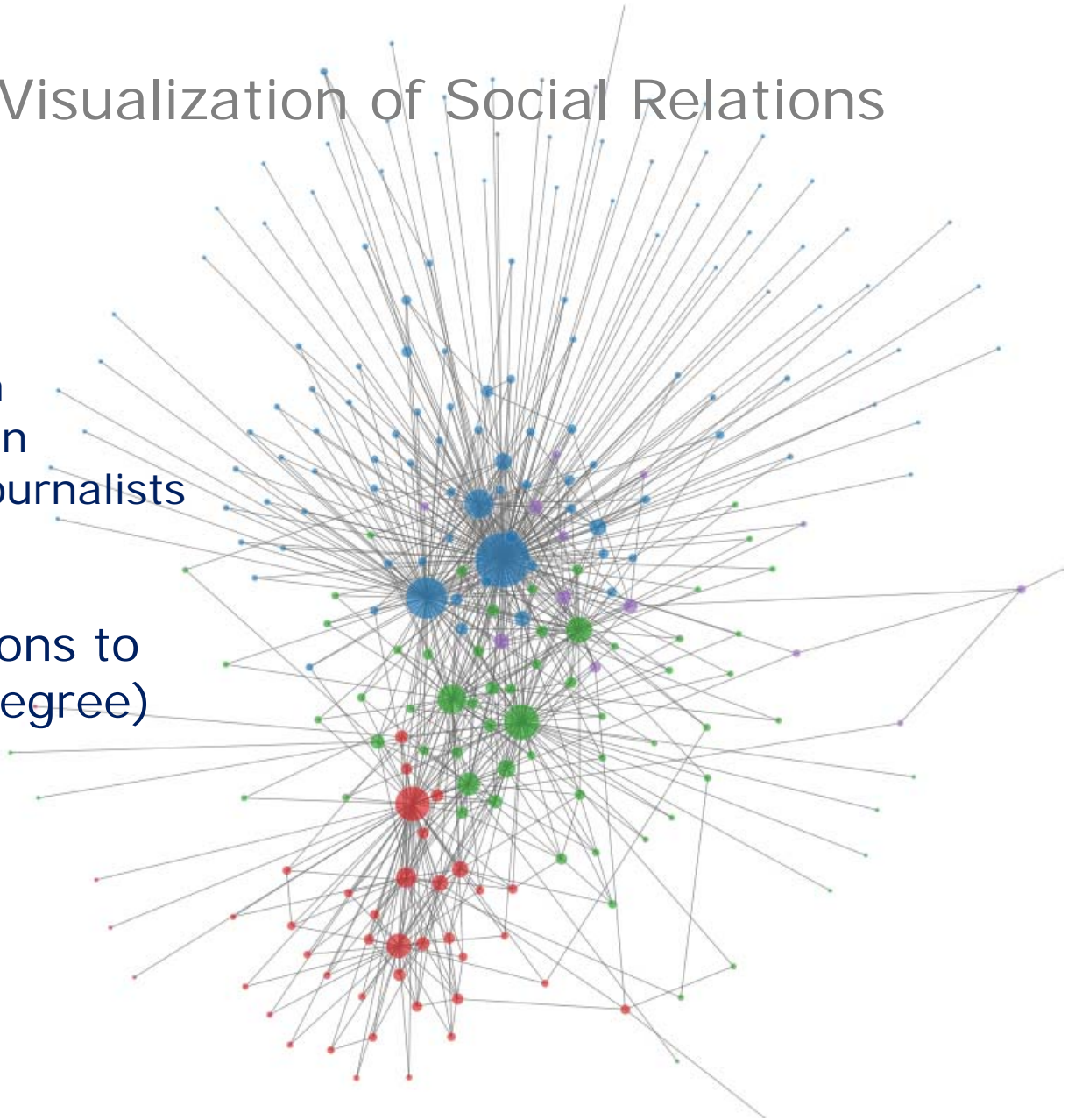
Soccer #fcbsvw

Groups

- FC Bayern Munich
- SV Werder Bremen
- News agencies, journalists

Number of connections to other users (node degree)

- 1
- 50
- 100
- 150



4 Analysis and Visualization of Social Relations

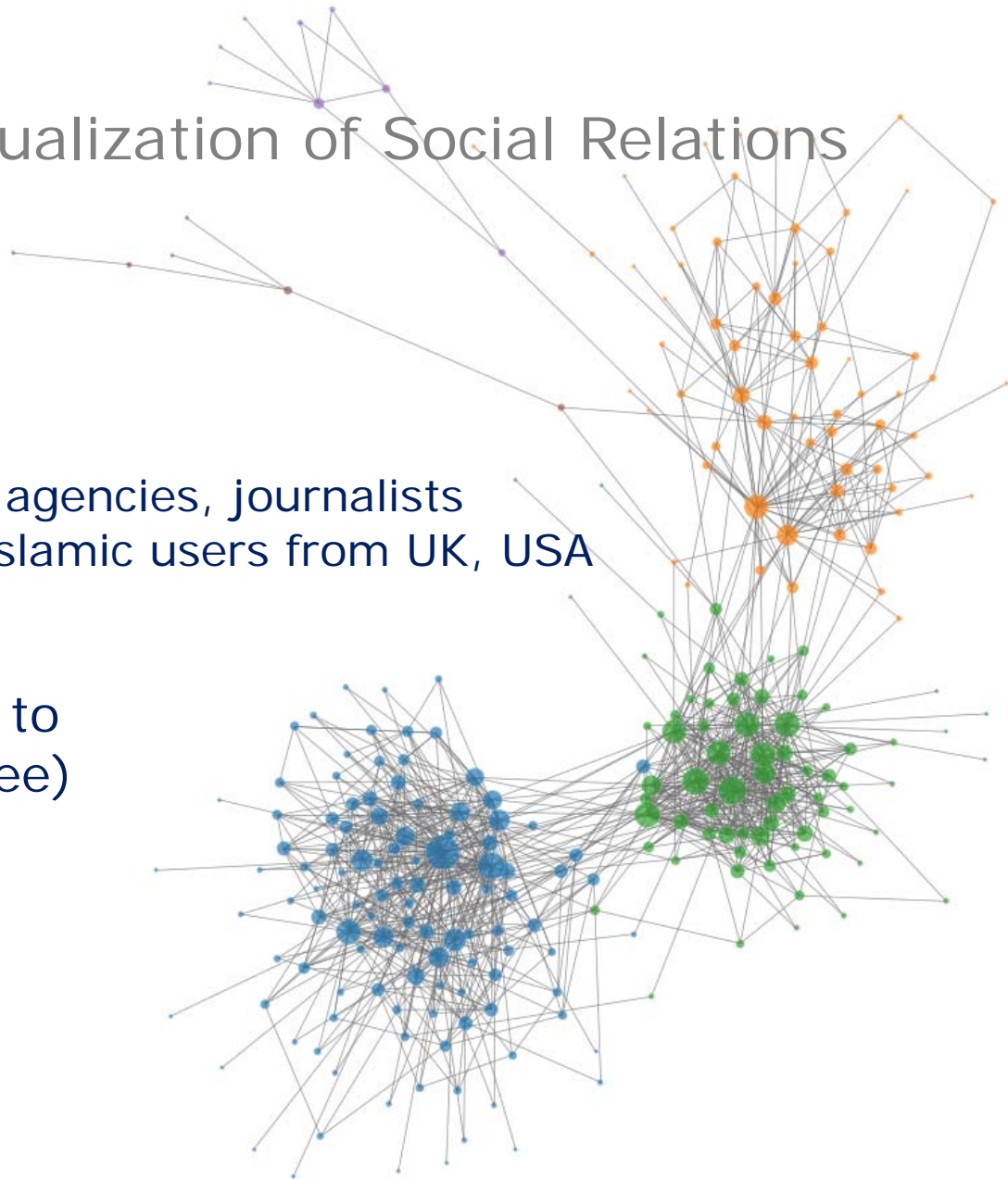
Politics #pegida

Groups

- Pro Pegida
- Against Pegida, news agencies, journalists
- Anti-European, Anti-Islamic users from UK, USA

Number of connections to other users (node degree)

- 1
- 50
- 100



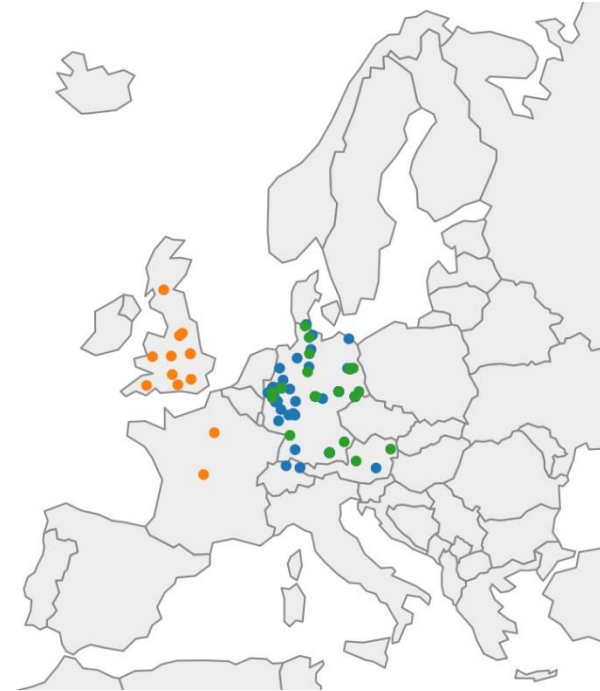
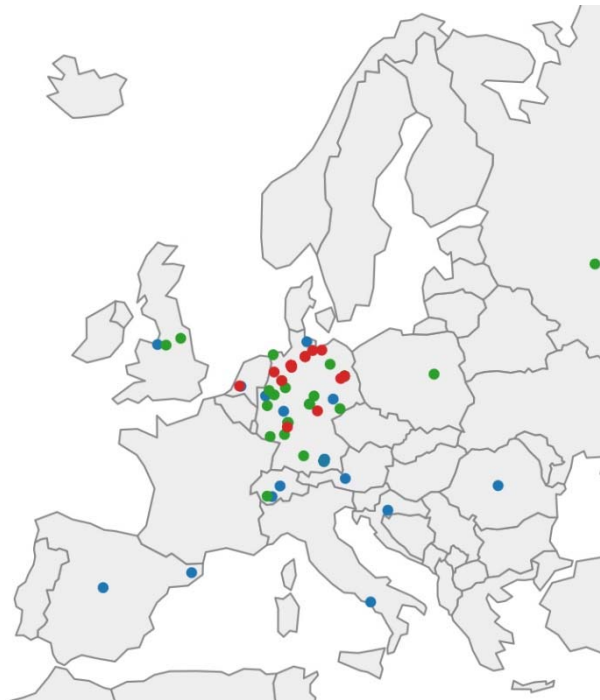
5 Analysis, Visualization of Geospatial Distribution

- Microblogging content often contains geospatial information
 - from profile of the user
 - from the tweet when the user optionally choose to provide location information
- In the two data sets none of the tweets had an explicit coordinate of the location
- The location information of the user profile were used

5 Analysis, Visualization of Geospatial Distribution

Soccer #fcbsvw

Politics #pegida



Users:

- 66% with location entry in profile
- 55% geocoded location entry
- 21% geolocated tweet in timeline

Users:

- 66% with location entry in profile
- 55% geocoded location entry
- 15% geolocated tweet in timeline

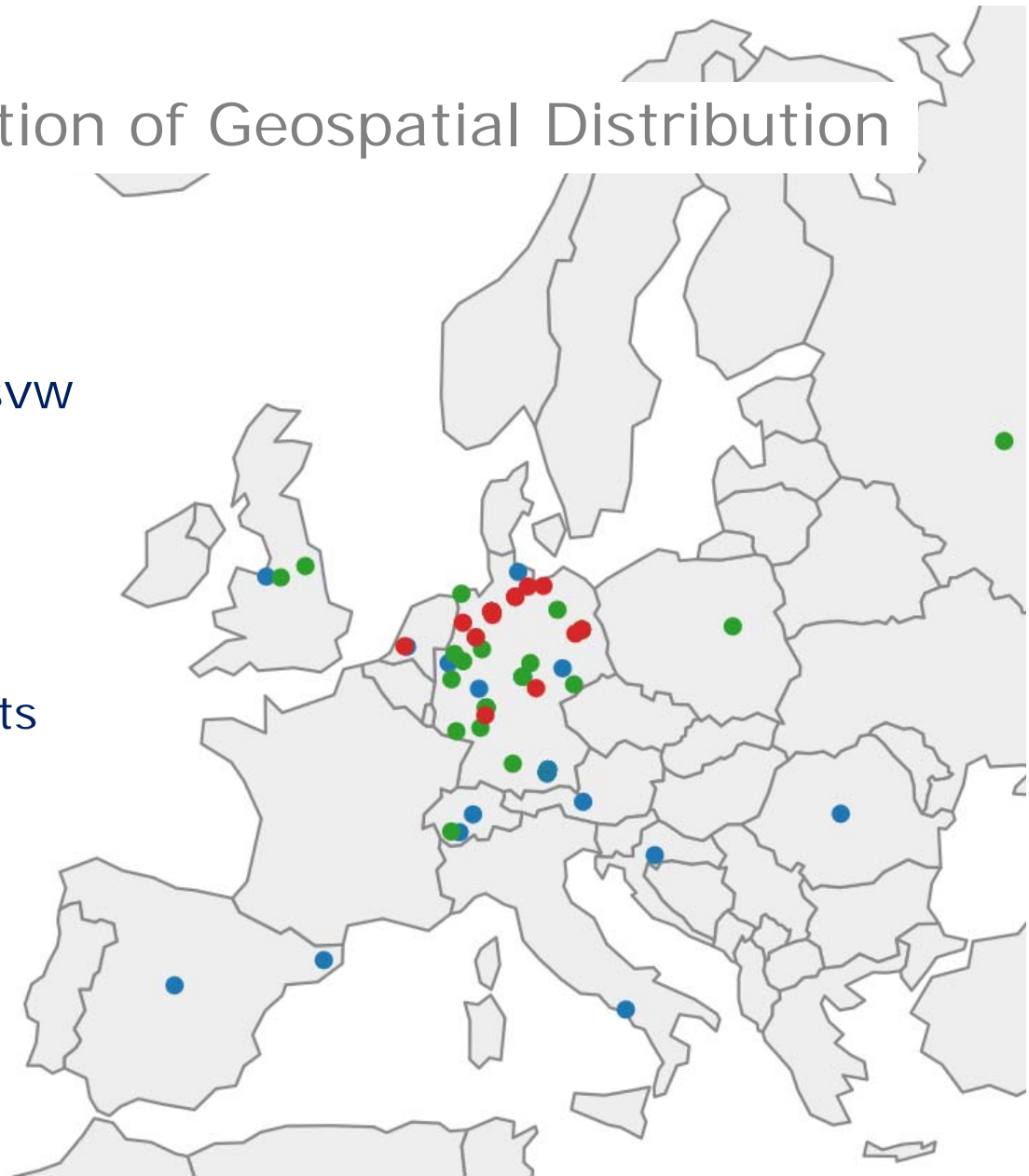
5 Analysis, Visualization of Geospatial Distribution

Geospatial distribution
of the users in Europe

Data set 1: Soccer #fcbsvw

Groups

- FC Bayern Munich
- SV Werder Bremen
- News agencies, journalists



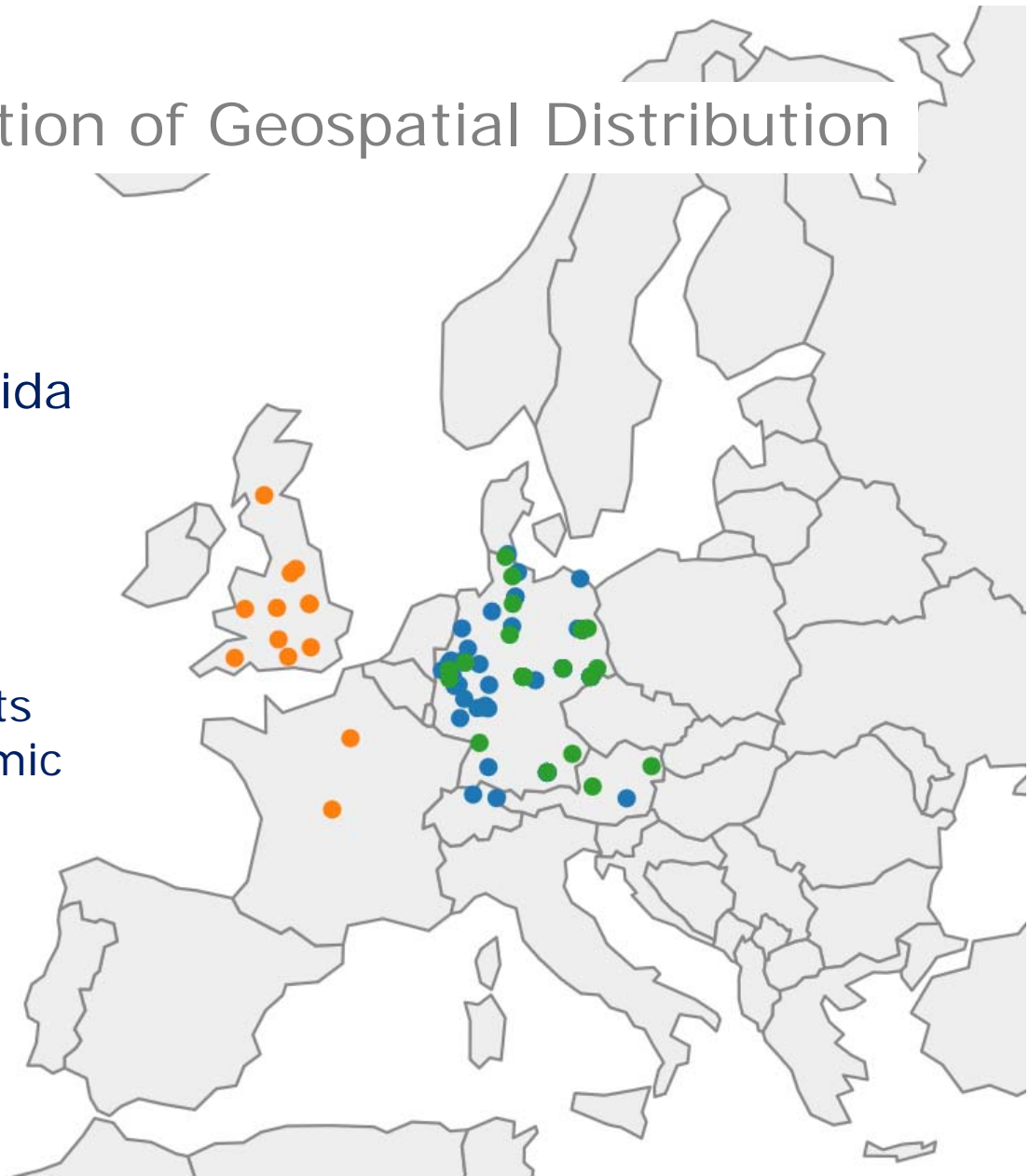
5 Analysis, Visualization of Geospatial Distribution

Geospatial distribution
of the users in Europe

Data set 2: Politics #pegida

Groups

- Pro Pegida
- Against Pegida,
news agencies, journalists
- Anti-European, Anti-Islamic
users from UK, USA





»Knowledge builds bridges.«