Typology, transformations, and clustering of spatio-temporal data

Lecture given by
prof. Gennady Andrienko and
prof. Natalia Andrienko
Content

• Types of spatio-temporal data: spatial events, spatial time series, and trajectories of moving objects.

• Transformations between the data types.

• Partition-based and density-based clustering.

• Two possible perspectives for looking at spatial time series and two complementary ways of applying partition-based clustering to them.

• Real-time detection of event concentrations
Types of spatio-temporal data

Data with spatial and temporal components
Spatial events

• A **spatial event** is a physical or abstract entity that appears at a certain time at a certain location in space.
  • **Instant event**: appears but does not exist any longer
    • or the existence time is negligibly small or out of interest for analysis
  • **Durable event**: exists during some time interval

• Examples
  • Instant: lightning flash, photo taken, tweet posted
  • Durable: election, stormy weather, New Year celebration

• Data structure:  <event identifier*, spatial position, time of appearing, time of disappearing or duration of existence**, any attributes>
  * May not be given explicitly; it is assumed that each data record describes one event
  ** Need to be specified only for durable events
Example: posted tweets

<table>
<thead>
<tr>
<th>identifier</th>
<th>LONGITUDE</th>
<th>LATITUDE</th>
<th>MESSAGE DATE</th>
<th>USERID</th>
<th>USERSCREENNAME</th>
<th>MESSAGE TEXT</th>
</tr>
</thead>
<tbody>
<tr>
<td>3936903236</td>
<td>13.121572494506836</td>
<td>52.38300323468328</td>
<td>25/10/2013 12:47:42</td>
<td>5755692</td>
<td>matesl</td>
<td>Recruiting Session at</td>
</tr>
<tr>
<td>3936903880</td>
<td>13.382525444030762</td>
<td>52.5311126709844</td>
<td>25/10/2013 12:47:57</td>
<td>218215137</td>
<td>buschensemble</td>
<td>@MusiciansBFund @F</td>
</tr>
<tr>
<td>3936905674</td>
<td>13.384698867797852</td>
<td>52.52968215942383</td>
<td>25/10/2013 12:48:40</td>
<td>14331417</td>
<td>WolfgangBremer</td>
<td>I'm at Bonfini (Berlin) h</td>
</tr>
<tr>
<td>3936906361</td>
<td>13.119993209838867</td>
<td>52.3819046025078</td>
<td>25/10/2013 12:48:56</td>
<td>160874621</td>
<td>_BB_RADIO_MUSIC</td>
<td>#nowplaying #stromae</td>
</tr>
<tr>
<td>3936907307</td>
<td>13.11646842956543</td>
<td>52.38713383669922</td>
<td>25/10/2013 12:49:19</td>
<td>161262801</td>
<td>RadioTeddyMusic</td>
<td>#nowplaying #silly ~ Si</td>
</tr>
<tr>
<td>3936907521</td>
<td>13.38502972785645</td>
<td>52.45845413260808</td>
<td>25/10/2013 12:49:24</td>
<td>88318861</td>
<td>ApfelVonSodom</td>
<td>@a_rabella eue te amo</td>
</tr>
<tr>
<td>3936908998</td>
<td>13.84437561035156</td>
<td>52.5071334838672</td>
<td>25/10/2013 12:49:59</td>
<td>173684513</td>
<td>a_abella</td>
<td>Just posted a photo</td>
</tr>
<tr>
<td>3936909835</td>
<td>13.45739459991455</td>
<td>52.51235580444336</td>
<td>25/10/2013 12:50:19</td>
<td>14776696</td>
<td>ninalemos</td>
<td>Sabe o que ando pens</td>
</tr>
<tr>
<td>3936911113</td>
<td>13.38885784149171</td>
<td>52.526232522563008</td>
<td>25/10/2013 12:50:54</td>
<td>24479611</td>
<td>sebastianwaters</td>
<td>new Lunch Menu (@ R</td>
</tr>
<tr>
<td>3936911396</td>
<td>13.23953914842334</td>
<td>52.514625549318406</td>
<td>25/10/2013 12:50:56</td>
<td>182001862</td>
<td>SINNBUERO</td>
<td>be berlin. #relax @ Oly</td>
</tr>
<tr>
<td>3936913661</td>
<td>13.35097480773926</td>
<td>52.59198760966328</td>
<td>25/10/2013 12:51:50</td>
<td>1889365722</td>
<td>KingslandRoadDE</td>
<td>@KingslandRd hi, I lov</td>
</tr>
<tr>
<td>3936916780</td>
<td>13.460264205932617</td>
<td>52.4754753112793</td>
<td>25/10/2013 12:53:05</td>
<td>1929914154</td>
<td>HOWLmariagluck</td>
<td>Thank you so much for</td>
</tr>
<tr>
<td>3936919251</td>
<td>13.412495613098145</td>
<td>52.50828552246094</td>
<td>25/10/2013 12:54:18</td>
<td>176009999</td>
<td>ngamzegungor</td>
<td>@IrmaKTanriover AMAI</td>
</tr>
<tr>
<td>3936921491</td>
<td>13.4556198287963687</td>
<td>52.53550720214844</td>
<td>25/10/2013 12:54:57</td>
<td>1077825853</td>
<td>thomasmatzka</td>
<td>Oh mein Gott!!! Der @C</td>
</tr>
<tr>
<td>3936925416</td>
<td>13.423683166503906</td>
<td>52.491004943847656</td>
<td>25/10/2013 12:56:31</td>
<td>541618147</td>
<td>EniseCoruh</td>
<td>Özledi im lezzet :) @ La</td>
</tr>
<tr>
<td>3936925533</td>
<td>13.451854705810547</td>
<td>52.5484962463389</td>
<td>25/10/2013 12:56:33</td>
<td>137301856</td>
<td>mrmojorisin</td>
<td>Mi è passata la voglia</td>
</tr>
<tr>
<td>3936926831</td>
<td>13.376999855041504</td>
<td>52.5161018371582</td>
<td>25/10/2013 12:57:04</td>
<td>1336218432</td>
<td>trendinaliaDE</td>
<td>#Snowden nur ein Tre</td>
</tr>
<tr>
<td>3936927118</td>
<td>13.394739151000977</td>
<td>52.47782516479492</td>
<td>25/10/2013 12:57:13</td>
<td>21513942</td>
<td>ralphcochrane</td>
<td>Getting ready for the cr</td>
</tr>
<tr>
<td>3936928716</td>
<td>13.388001441955566</td>
<td>52.49345016479492</td>
<td>25/10/2013 12:57:49</td>
<td>20170781</td>
<td>david_g_cooper</td>
<td>German bar staff need</td>
</tr>
<tr>
<td>3936929248</td>
<td>13.29820671081543</td>
<td>52.45787429609579</td>
<td>25/10/2013 12:58:02</td>
<td>92283308</td>
<td>x_elly_</td>
<td>So dreckig wie heute g</td>
</tr>
<tr>
<td>3936929708</td>
<td>13.414163357849121</td>
<td>52.5078659507617</td>
<td>25/10/2013 12:58:13</td>
<td>175051281</td>
<td>betuel07</td>
<td>I'm at Bahçe ehir Unive</td>
</tr>
<tr>
<td>3936931181</td>
<td>13.11768913269043</td>
<td>52.384178161621094</td>
<td>25/10/2013 12:58:48</td>
<td>119337117</td>
<td>quangVFX</td>
<td>I'm at Animago Confer</td>
</tr>
</tbody>
</table>
Spatial positions of the tweet posting events

Spatio-temporal positions of the events

Notes:
1. Events may have non-zero spatial extents (not necessarily points).
2. Durable events can be represented in a space-time cube by vertical bars or prisms.

The idea of **space-time cube** comes from:
“What about people in regional science?”
Papers of the Regional Science Association; 24:7-21.
Trajectories of moving objects

• A **trajectory** is a chronologically ordered sequence of time-referenced spatial positions of a moving object

• Examples: GPS tracks of vehicles or animals

• Data structure: <object identifier and/or trip identifier, time stamp, spatial position, *any attributes*>
### Example: trajectories of vehicles

<table>
<thead>
<tr>
<th>Trip identifier</th>
<th>Time stamp</th>
<th>Spatial position</th>
<th>Object identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>longitude</td>
<td>latitude</td>
</tr>
<tr>
<td>339_900</td>
<td>06/03/2007 19:17:31</td>
<td>-0.10423000156879425</td>
<td>51.50750732421875</td>
</tr>
<tr>
<td>339_901</td>
<td>06/03/2007 19:17:39</td>
<td>-0.1042733414554596</td>
<td>51.50740814208944</td>
</tr>
<tr>
<td>339_902</td>
<td>06/03/2007 19:17:49</td>
<td>-0.10411166399717331</td>
<td>51.50752258300781</td>
</tr>
<tr>
<td>339_903</td>
<td>06/03/2007 19:18:02</td>
<td>-0.1044683307492865</td>
<td>51.5071220397942</td>
</tr>
<tr>
<td>339_904</td>
<td>06/03/2007 19:18:10</td>
<td>-0.1043633297085762</td>
<td>51.50709533691406</td>
</tr>
<tr>
<td>339_905</td>
<td>06/03/2007 19:18:51</td>
<td>-0.10951333492994308</td>
<td>51.50588226318359</td>
</tr>
<tr>
<td>339_906</td>
<td>06/03/2007 19:19:13</td>
<td>-0.11245166510343552</td>
<td>51.5056268784668</td>
</tr>
<tr>
<td>339_907</td>
<td>06/03/2007 19:19:23</td>
<td>-0.11319166421890259</td>
<td>51.5044822692871</td>
</tr>
<tr>
<td>339_908</td>
<td>06/03/2007 19:19:34</td>
<td>-0.1137833297252665</td>
<td>51.5042266845703</td>
</tr>
<tr>
<td>339_909</td>
<td>06/03/2007 19:19:44</td>
<td>-0.1153983324766159</td>
<td>51.50351333618164</td>
</tr>
<tr>
<td>345_0</td>
<td>06/03/2007 08:31:06</td>
<td>0.0583166666328907</td>
<td>51.52149200439453</td>
</tr>
<tr>
<td>345_1</td>
<td>06/03/2007 08:31:17</td>
<td>0.058694999665021896</td>
<td>51.52083206176758</td>
</tr>
<tr>
<td>345_2</td>
<td>06/03/2007 08:31:30</td>
<td>0.0591333322270012</td>
<td>51.52022171020508</td>
</tr>
<tr>
<td>345_3</td>
<td>06/03/2007 08:31:38</td>
<td>0.05857166667411072</td>
<td>51.51984405517578</td>
</tr>
<tr>
<td>345_4</td>
<td>06/03/2007 08:31:48</td>
<td>0.05670500174164772</td>
<td>51.519508361816406</td>
</tr>
<tr>
<td>345_5</td>
<td>06/03/2007 08:31:58</td>
<td>0.05439166724681854</td>
<td>51.51934814453125</td>
</tr>
<tr>
<td>345_6</td>
<td>06/03/2007 08:32:09</td>
<td>0.052186667919158936</td>
<td>51.519493103023734</td>
</tr>
<tr>
<td>345_7</td>
<td>06/03/2007 08:32:19</td>
<td>0.04961499443292626</td>
<td>51.51976776123047</td>
</tr>
<tr>
<td>345_8</td>
<td>06/03/2007 08:32:31</td>
<td>0.0467749830722809</td>
<td>51.52006149291992</td>
</tr>
<tr>
<td>345_9</td>
<td>06/03/2007 08:32:39</td>
<td>0.044165000319480896</td>
<td>51.52008056640625</td>
</tr>
<tr>
<td>345_10</td>
<td>06/03/2007 08:32:49</td>
<td>0.0417400024795532</td>
<td>51.519737243652344</td>
</tr>
<tr>
<td>345_11</td>
<td>06/03/2007 08:33:02</td>
<td>0.0400283336394043</td>
<td>51.519317626953125</td>
</tr>
</tbody>
</table>
Spatial footprints of multiple trajectories
Spatial (= spatially referenced) time series

- A **time series** is a chronologically ordered sequence of *data items* that refer to different moments or intervals in time
  - Time series of attribute values
    - E.g., measured values of weather parameters at different times
    - Time series of satellite images
  - A **spatial time series** is a set of time series of attribute values, where each time series refers to a distinct spatial location (point or region in space) or a spatial object.
### Example: Crime Statistics

#### Attributes

<table>
<thead>
<tr>
<th>Year</th>
<th>Id</th>
<th>State</th>
<th>Population</th>
<th>Index Offenses</th>
<th>Violent Offenses</th>
<th>Murder</th>
<th>Forcible Rape</th>
<th>Robbery</th>
<th>Aggravated Assault</th>
<th>Property Crime</th>
<th>Burglary</th>
<th>Larceny-Theft</th>
<th>Motor Vehicle Theft</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960</td>
<td>1</td>
<td>Alabama</td>
<td>3256740</td>
<td>35920</td>
<td>6097</td>
<td>406</td>
<td>281</td>
<td>898</td>
<td>4512</td>
<td>33823</td>
<td>11626</td>
<td>19344</td>
<td>2853</td>
</tr>
<tr>
<td>1960</td>
<td>2</td>
<td>Alaska</td>
<td>225167</td>
<td>3730</td>
<td>236</td>
<td>23</td>
<td>47</td>
<td>64</td>
<td>102</td>
<td>3494</td>
<td>751</td>
<td>2195</td>
<td>548</td>
</tr>
<tr>
<td>1960</td>
<td>4</td>
<td>Arizona</td>
<td>130216</td>
<td>35243</td>
<td>2704</td>
<td>78</td>
<td>209</td>
<td>706</td>
<td>1711</td>
<td>35339</td>
<td>8926</td>
<td>23207</td>
<td>4406</td>
</tr>
<tr>
<td>1960</td>
<td>5</td>
<td>Arkansas</td>
<td>1785272</td>
<td>18472</td>
<td>1924</td>
<td>152</td>
<td>159</td>
<td>443</td>
<td>1170</td>
<td>16548</td>
<td>5899</td>
<td>10250</td>
<td>899</td>
</tr>
<tr>
<td>1960</td>
<td>6</td>
<td>California</td>
<td>15717204</td>
<td>546609</td>
<td>37556</td>
<td>616</td>
<td>2859</td>
<td>15287</td>
<td>18796</td>
<td>5038511</td>
<td>143102</td>
<td>311956</td>
<td>53453</td>
</tr>
<tr>
<td>1960</td>
<td>8</td>
<td>Colorado</td>
<td>1753947</td>
<td>38103</td>
<td>2408</td>
<td>73</td>
<td>229</td>
<td>1362</td>
<td>744</td>
<td>35995</td>
<td>9996</td>
<td>21949</td>
<td>3750</td>
</tr>
<tr>
<td>1960</td>
<td>9</td>
<td>Connecticut</td>
<td>2535234</td>
<td>29231</td>
<td>928</td>
<td>41</td>
<td>103</td>
<td>236</td>
<td>548</td>
<td>23393</td>
<td>8452</td>
<td>16653</td>
<td>3288</td>
</tr>
<tr>
<td>1960</td>
<td>10</td>
<td>Delaware</td>
<td>445292</td>
<td>9642</td>
<td>375</td>
<td>33</td>
<td>41</td>
<td>157</td>
<td>144</td>
<td>9267</td>
<td>2661</td>
<td>5857</td>
<td>739</td>
</tr>
<tr>
<td>1960</td>
<td>11</td>
<td>District of Columbia</td>
<td>793955</td>
<td>20725</td>
<td>4230</td>
<td>81</td>
<td>111</td>
<td>1072</td>
<td>2955</td>
<td>16495</td>
<td>4587</td>
<td>9905</td>
<td>2003</td>
</tr>
<tr>
<td>1960</td>
<td>12</td>
<td>Florida</td>
<td>4951560</td>
<td>133519</td>
<td>11061</td>
<td>527</td>
<td>403</td>
<td>4005</td>
<td>6126</td>
<td>122858</td>
<td>39966</td>
<td>73030</td>
<td>9289</td>
</tr>
</tbody>
</table>

#### Additional Data

- **Reference 1**: Time
- **Reference 2**: Place

....

1972: 54 West Virginia
1972: 55 Wisconsin
1972: 56 Wyoming
1973: 4 Arizona
1973: 5 Arkansas
1973: 6 California
1973: 8 Colorado
1973: 9 Connecticut
1974: 44 Rhode Island
1975: 45 South Carolina
1975: 46 South Dakota
1975: 47 Tennessee
1975: 48 Texas
1975: 49 Utah
1975: 50 Vermont
1975: 51 Virginia
1975: 52 Washington
1975: 55 Wisconsin
1975: 56 Wyoming
References are not always in table columns

Reference 1: place

Reference 2: time

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbey</td>
<td>278</td>
<td>303</td>
<td>279</td>
<td>300</td>
<td>248</td>
<td>255</td>
<td>209</td>
<td>215</td>
<td>221</td>
</tr>
<tr>
<td>Albion</td>
<td>79</td>
<td>84</td>
<td>86</td>
<td>81</td>
<td>90</td>
<td>97</td>
<td>89</td>
<td>95</td>
<td>97</td>
</tr>
<tr>
<td>Becontree</td>
<td>96</td>
<td>80</td>
<td>105</td>
<td>108</td>
<td>112</td>
<td>110</td>
<td>101</td>
<td>102</td>
<td>103</td>
</tr>
<tr>
<td>Chadwell Heath</td>
<td>118</td>
<td>133</td>
<td>157</td>
<td>157</td>
<td>153</td>
<td>136</td>
<td>127</td>
<td>129</td>
<td>112</td>
</tr>
<tr>
<td>Eastbrook</td>
<td>75</td>
<td>86</td>
<td>77</td>
<td>75</td>
<td>31</td>
<td>82</td>
<td>87</td>
<td>85</td>
<td>84</td>
</tr>
<tr>
<td>Eastbury</td>
<td>100</td>
<td>98</td>
<td>109</td>
<td>119</td>
<td>105</td>
<td>134</td>
<td>121</td>
<td>106</td>
<td>97</td>
</tr>
<tr>
<td>Goscoigne</td>
<td>277</td>
<td>239</td>
<td>220</td>
<td>190</td>
<td>170</td>
<td>162</td>
<td>133</td>
<td>123</td>
<td>107</td>
</tr>
<tr>
<td>Gresbrook</td>
<td>75</td>
<td>75</td>
<td>99</td>
<td>103</td>
<td>99</td>
<td>86</td>
<td>88</td>
<td>89</td>
<td>93</td>
</tr>
<tr>
<td>Heath</td>
<td>104</td>
<td>107</td>
<td>104</td>
<td>117</td>
<td>115</td>
<td>128</td>
<td>119</td>
<td>107</td>
<td>93</td>
</tr>
<tr>
<td>Londobridge</td>
<td>76</td>
<td>67</td>
<td>82</td>
<td>93</td>
<td>71</td>
<td>65</td>
<td>77</td>
<td>73</td>
<td>75</td>
</tr>
<tr>
<td>Mayesbrook</td>
<td>90</td>
<td>74</td>
<td>72</td>
<td>92</td>
<td>95</td>
<td>97</td>
<td>92</td>
<td>89</td>
<td>94</td>
</tr>
<tr>
<td>Parishes</td>
<td>77</td>
<td>79</td>
<td>70</td>
<td>92</td>
<td>95</td>
<td>97</td>
<td>92</td>
<td>89</td>
<td>84</td>
</tr>
<tr>
<td>River</td>
<td>113</td>
<td>103</td>
<td>114</td>
<td>102</td>
<td>115</td>
<td>119</td>
<td>110</td>
<td>99</td>
<td>96</td>
</tr>
<tr>
<td>Thames</td>
<td>263</td>
<td>278</td>
<td>254</td>
<td>237</td>
<td>215</td>
<td>208</td>
<td>180</td>
<td>178</td>
<td>175</td>
</tr>
<tr>
<td>Valentia</td>
<td>81</td>
<td>76</td>
<td>83</td>
<td>87</td>
<td>98</td>
<td>89</td>
<td>76</td>
<td>96</td>
<td>74</td>
</tr>
<tr>
<td>Village</td>
<td>111</td>
<td>124</td>
<td>143</td>
<td>120</td>
<td>115</td>
<td>134</td>
<td>130</td>
<td>129</td>
<td>133</td>
</tr>
<tr>
<td>Whalebone</td>
<td>109</td>
<td>93</td>
<td>96</td>
<td>97</td>
<td>114</td>
<td>115</td>
<td>99</td>
<td>97</td>
<td>107</td>
</tr>
<tr>
<td>BrunswickPark</td>
<td>70</td>
<td>88</td>
<td>94</td>
<td>79</td>
<td>88</td>
<td>81</td>
<td>56</td>
<td>54</td>
<td>58</td>
</tr>
<tr>
<td>Burnt Oak</td>
<td>88</td>
<td>106</td>
<td>102</td>
<td>114</td>
<td>95</td>
<td>95</td>
<td>79</td>
<td>70</td>
<td>76</td>
</tr>
<tr>
<td>Childs Hill</td>
<td>119</td>
<td>125</td>
<td>129</td>
<td>142</td>
<td>138</td>
<td>127</td>
<td>118</td>
<td>106</td>
<td>106</td>
</tr>
<tr>
<td>Collindale</td>
<td>92</td>
<td>114</td>
<td>114</td>
<td>107</td>
<td>95</td>
<td>93</td>
<td>76</td>
<td>76</td>
<td>70</td>
</tr>
<tr>
<td>Coppets</td>
<td>99</td>
<td>104</td>
<td>115</td>
<td>113</td>
<td>100</td>
<td>68</td>
<td>76</td>
<td>75</td>
<td>91</td>
</tr>
<tr>
<td>East Barnet</td>
<td>61</td>
<td>62</td>
<td>80</td>
<td>88</td>
<td>82</td>
<td>61</td>
<td>66</td>
<td>65</td>
<td>81</td>
</tr>
<tr>
<td>East Finchley</td>
<td>85</td>
<td>87</td>
<td>95</td>
<td>92</td>
<td>81</td>
<td>84</td>
<td>59</td>
<td>67</td>
<td>81</td>
</tr>
<tr>
<td>Edgware</td>
<td>107</td>
<td>106</td>
<td>107</td>
<td>121</td>
<td>114</td>
<td>106</td>
<td>89</td>
<td>84</td>
<td>95</td>
</tr>
<tr>
<td>Finchley Church</td>
<td>63</td>
<td>59</td>
<td>58</td>
<td>54</td>
<td>52</td>
<td>56</td>
<td>56</td>
<td>51</td>
<td>53</td>
</tr>
<tr>
<td>Garden Suburb</td>
<td>97</td>
<td>83</td>
<td>105</td>
<td>105</td>
<td>70</td>
<td>72</td>
<td>74</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>Golders Green</td>
<td>94</td>
<td>96</td>
<td>101</td>
<td>112</td>
<td>96</td>
<td>96</td>
<td>79</td>
<td>70</td>
<td>83</td>
</tr>
<tr>
<td>Hale</td>
<td>56</td>
<td>63</td>
<td>65</td>
<td>70</td>
<td>72</td>
<td>70</td>
<td>54</td>
<td>54</td>
<td>55</td>
</tr>
<tr>
<td>Hendon</td>
<td>98</td>
<td>83</td>
<td>101</td>
<td>94</td>
<td>106</td>
<td>87</td>
<td>84</td>
<td>93</td>
<td>92</td>
</tr>
<tr>
<td>High Barnet</td>
<td>79</td>
<td>92</td>
<td>105</td>
<td>109</td>
<td>101</td>
<td>75</td>
<td>80</td>
<td>80</td>
<td>72</td>
</tr>
<tr>
<td>Mill Hill</td>
<td>79</td>
<td>90</td>
<td>103</td>
<td>93</td>
<td>85</td>
<td>90</td>
<td>84</td>
<td>74</td>
<td>71</td>
</tr>
<tr>
<td>Oakleigh</td>
<td>57</td>
<td>63</td>
<td>72</td>
<td>70</td>
<td>72</td>
<td>61</td>
<td>58</td>
<td>62</td>
<td>51</td>
</tr>
</tbody>
</table>
Slices of spatial time series

1) Space as a whole
   Selected time $t_k$
   Slice: spatial behaviour (= distribution of attribute values over space) at this time

2) Selected place
   Time as a whole
   Slice: temporal behaviour (= variation of attribute values over time) in this place
Two complementary views of spatial time series

1. As a spatial distribution of local time series: a set of time series of attribute values in different locations

2. As time-varying spatial situations: sequence of spatial distributions of attribute values at different times

These views require different visualisation and analysis techniques.
Transformations of spatio-temporal data
Transformations of spatio-temporal data

- Spatial events
  - integrate
  - disintegrate, extract

- Trajectories
  - divide
  - aggregate

- Spatial time series (place-based)
  - extract
  - aggregate

- Spatial time series (link-based)

- Local time series
  - aspects (views)

- Spatial situations
Spatial events $\rightarrow$ trajectories

Tweet events of one Twitter user

Trajectory of the Twitter user
Spatial events → trajectories

Tweet events of multiple users
Quasi-continuous vs. episodic trajectories

Quasi-continuous:
• Small spatial and temporal distances between consecutive positions
• Spatially smooth
• Permit interpolation between recorded positions.

Episodic:
• May have large spatial and temporal gaps between consecutive positions
• Spatially abrupt
• No valid interpolation between recorded positions is possible.
Trajectories $\rightarrow$ spatial events (of interest)

E.g., stops for at least 5 minutes
Spatio-temporal aggregation of events

- Spatial events and their thematic attributes can be aggregated spatially by spatial compartments and time intervals.
  - Partition the underlying space extent into suitable compartments (areas)
    - Create a regular or irregular grid
    - Use a pre-existing division (e.g., administrative)
  - Partition the time span of the data into intervals
  - For each compartment and time interval:
    - Count the events; possibly, normalize by compartment areas, resident population, ...
    - Compute statistics of values of thematic attributes

- Resulting data type: **spatial time series** of
  - event counts, densities, counts per capita, ...
  - statistical summaries of thematic attributes: mean, median, mode, minimum, maximum, quantiles, ...
Voronoi tessellation (a.k.a. Voronoi diagram)

*Used for building irregular grids*

- The partitioning of a plane with N points into convex polygons (*cells*), such that
  - each polygon contains exactly one generating point
  - every point in a given polygon is closer to its generating point than to any other.
- The generating points are also called *seeds*.
- A Voronoi diagram is also known as a Dirichlet tessellation.
- The cells are called Dirichlet regions, Thiessen polytopes, or Voronoi polygons.
Data-driven tessellation

**Step 1:** apply a special algorithm for grouping of points based on their spatial proximity.

The main idea: put the points into circles with a given maximal radius $R$.

In this example, $R = 1000$ m.

**Step 2:** use the centres of the point groups as the generating seeds for the Voronoi tessellation.
Spatial events → spatial time series
Spatio-temporal aggregation of trajectories

- Partition the underlying space extent into suitable compartments (areas)
- Partition the time span of the data into intervals

1. For each compartment $C$ and time interval $\Delta t$, count the number of visits that occurred during $\Delta t$ and the number of distinct visitors → place-based time series (time series of presence)

2. For each pair of compartments $C_1$ and $C_2$ and time interval $\Delta t$, count the number of moves (transitions) from $C_1$ to $C_2$ that occurred during $\Delta t$ and the number of distinct objects that moved → link-based time series (time series of aggregate moves, called flows)
Trajectories → place-based time series (presence)

The map shows the counts of distinct visitors transformed to the differences from the means.
Trajectories → link-based time series (flows)
The flow symbols

Arrow: shows the movement direction; the width can represent a numeric attribute, such as the number of moving objects or the count of moves.

Problem: it is hard to represent movements in two opposite directions.
Solution: use halves of arrows!

Movement to the right

Movement to the left

Easy to put together:

Can show asymmetric amounts of movement
Another variant of flow symbols
Flows obtained from episodic trajectories often intersect in space. They are hard to represent on a map in a readable manner. This does not mean, however, that such flows cannot be analysed.
Spatial time series → spatial events

Events in time series: peaks, drops, trend change, ...
Events of interest can be **extracted** from time series.
Events extracted from spatial time series are spatial events.
Transformations of spatio-temporal data

Transformations enable multi-perspective analyses of spatio-temporal data.

- Spatial events
  - integrate
  - disintegrate, extract

- Trajectories
  - divide

- Spatial time series (place-based)
  - aggregate
  - extract

- Spatial time series (link-based)
  - aggregate

- Local time series
  - aspects (views)

- Spatial situations
Introduction to clustering

Partition-based and density-based clustering
What is clustering?

• Loose definition: clustering is the process of organising objects into groups whose members are close or similar in some way.

• A cluster is a group of objects which are “similar” or “close” between them and are “dissimilar” or “distant” to the objects belonging to other clusters.
Clusters of similar time series

Clusters of spatially close spatial objects

Clusters of similar multi-attribute value combinations

Clusters of spatially and temporally close spatial events
Role of clustering in visual analytics

• Grouping of similar or close items plays an essential role in VA
  • as a tool supporting abstraction: elements $\rightarrow$ subsets;
    the subsets may be considered as wholes
  • as a tool to manage large data volumes
  • as a tool to find specific features of interest, e.g., event concentrations
  • as a tool to deal with multiple attributes and multiple time series, which are hard to visualise
Two major types of clustering

- **Partition-based clustering (PBC):** divide items into groups so that items within a group are similar (close) and items from different groups are less similar (more distant)
  - Examples: k-means, SOM (self-organizing map)
  - Property of the result: each item belongs to some group

- **Density-based clustering (DBC):** find groups of highly similar (close) items and separate from them items that are less similar (more distant) to others
  - Examples: DBScan, OPTICS
  - Properties of the results: some items belong to groups, other items remain ungrouped and are treated as “noise”
Example: DBC (left) and PBC (right) of points according to their spatial positions
Use of the two types of clustering

• **Partition-based:**
  - Typically applied to multiple thematic (non-spatial) attributes or to **time series** of thematic attributes
  - Objective: divide objects into groups such that objects within a group have similar attribute values and differ from the objects in the other groups

• **Density-based:**
  - Typically applied to spatial and temporal attributes of spatial or spatio-temporal objects
  - Objective: find concentrations of objects in space or in space and time (i.e., groups of objects with close spatial locations and existence times)
    - concentrations of objects may have special meanings; e.g., spatio-temporal cluster of low speed events ⇒ traffic jam
Partition-based clustering of spatial time series

- as local time series
- as spatial situations
PBC of local time series (e.g., k-means)

Result: clusters of places
PBC of spatial situations (e.g., k-means)

Result: clusters of times
Two-way PBC of spatial time series

PBC can be applied to table rows or to columns
Problem: what value of k to choose?

*Generally: for any computational tool, what parameter settings to choose?*

- Typically not known in advance
- Computation results (such as clusters) need to be properly visualised and examined
  - Clustering results are often represented by colour-coding, which is applied to different visual objects, depending on the structure of the input data
- The analyst needs to run the tool with different settings and see how the results change
- The analyst then selects the settings bringing the “best” results:
  - easy to interpret (e.g., understandable spatial patterns)
  - internal variance within the clusters is sufficiently low
  - fit to the purpose (e.g., the intended analysis scale may require coarser or finer division)
Density-based clustering
concept and parameters
Density-based clustering (DBC)

**Goal: find dense groups of close or similar objects**

- For a given object $o$, the objects whose distances from $o$ are within a chosen distance threshold (radius) $R$ are called *neighbours* of the object $o$.

- An object is treated as a *core* object of a cluster if it has at least $N$ neighbours.

- To make a cluster:
  1) some core object with all its neighbours is taken;
  2) for each core object already included in the cluster, all its neighbours are also added to the cluster (if not added yet).

- Some objects may remain out of any cluster (when they have not enough neighbours and do not belong to the neighbourhood of any core object). These objects are treated as “*noise*”.
Density-based clustering

Parameters

• For DBC, the user needs to specify the neighbourhood radius (distance threshold) $R$ and the minimum number of neighbours $N$.

⇒ The use of DBC requires an understandable definition of distance between objects
  • e.g., spatial distance or spatio-temporal distance.

• Results of DBC greatly depend on the parameter choice.

• Visualisation and interactive exploration help the analyst to find suitable values for $R$ and $N$ that lead to good results.

Grey: “noise”
Exploring the impact of the DBC parameters

*Example: spatial clusters of point objects*

R = 500m; N = 20

R = 250m; N = 20

R = 250m; N = 10

R = 100m; N = 10
Good parameter settings?

- **R = 500m; N = 20**
  - Some clusters are too loose and too extended in space.
  - 68 classes in total

- **R = 250m; N = 20**
  - Some clusters are still too loose.
  - 129 classes in total

- **R = 100m; N = 10**
  - Clusters are compact (but quite numerous).
  - 450 classes in total

- **R = 100m; N = 20**
  - Clusters are compact and less numerous, but too many objects go to the noise.
  - 270 classes in total
Density-based clustering of spatial events
by positions in space and time
DBC by spatio-temporal distances

*Used for finding spatio-temporal concentrations of spatial events*

- For any two objects, there is a distance in space $d_{\text{space}}$ and a distance in time $d_{\text{time}}$.

- To cluster the objects by their spatio-temporal proximity, the analyst may choose two neighbourhood radii $R_{\text{space}}$ and $R_{\text{time}}$.
  - e.g., $R_{\text{space}} = 300$ m and $R_{\text{time}} = 30$ minutes.

- However, the clustering algorithm requires a single distance and a single radius.

  $\Rightarrow$ Spatial and temporal distances need to be combined together

  - e.g., $d = \max(d_{\text{space}}/R_{\text{space}}, d_{\text{time}}/R_{\text{time}}) \times R_{\text{space}}$
Example: spatio-temporal clusters of tweet events

\[ R_{\text{space}} = 250 \text{ m}; \quad R_{\text{time}} = 15 \text{ minutes}; \quad 8,476 \text{ events (2 days); 78 clusters with size } \geq 10; \quad \text{largest cluster size: 370 events; noise: 5,897 events} \]
Density-based clustering: a summary

• Goal: find groups of highly similar (close) items and separate from them items that are less similar (more distant) to others.

• DBC is often applied to spatial and spatio-temporal objects
  • to find spatial and spatio-temporal concentrations of objects;
  • to find groups of objects with similar spatial or spatio-temporal properties

• Parameters:
  • distance threshold (neighbourhood radius) \( R \)
  • minimal number of neighbours of a cluster core object \( N \)

• The analyst needs to set a meaningful distance threshold
  \( \Rightarrow \) Well understandable distances between items must exist
    • Spatial distance, temporal distance, difference between directions, ...
Distance functions in DBC

- Elementary distances: spatial, temporal, difference of values of a single thematic attribute

- It may be necessary to group objects on the basis of two or more elementary distances, e.g., spatial and temporal

  ⇒ A distance function integrating the elementary distances is needed

- General approach:
  1) Set a separate threshold for each elementary distance
  2) Transform the absolute elementary distances to relative w.r.t. the respective thresholds
  3) Combine the relative distances:
     - take their maximum or compute the Euclidean or Manhattan distance
Investigation of parameter impact

• The results of DBC greatly depend on the parameter settings (values of \(R\) and \(N\))

⇒ It is necessary to run the clustering tool multiple times with different parameter settings
  • Choose clear, easily interpretable results
  • Results from different runs may complement each other and contribute to better understanding

• Interactive visual interfaces are used for investigating the results of different runs.
Two major types of clustering: a reminder

- **Partition-based clustering**: divide items into groups so that items within a group are similar (close) and items from different groups are less similar (more distant)
  - Examples: k-means, self-organizing map
  - Property of the result: each item belongs to some group

- **Density-based clustering**: find groups of highly similar (close) items and separate from them items that are less similar (more distant) to others
  - Examples: DBScan, OPTICS
  - Properties of the results: some items belong to groups, other items remain ungrouped and are treated as “noise”
Density-based clustering of streaming events

Real-time detection and tracking of spatio-temporal concentrations of events
Problem setting

• Scenario:
  • Spatial events are registered in real time → stream of event data
  • Any individual event is not significant but spatio-temporal concentrations (clusters) of events are.
  • Monitoring task: detect emerging clusters of spatial events and trace their further evolution
• Our goal: support the observer
  • Automatically detect event clusters in the stream as soon as they emerge
  • Visually present the detected clusters and all their further changes to the observer
Motivating examples

Among geo-located tweets, some may mention extreme weather conditions (e.g., a storm). Spatio-temporal clusters of such tweets may indicate that certain regions are strongly affected by these conditions.

Spatio-temporal clusters of low speed events with similar movement directions may indicate traffic jams. Spatio-temporal clusters of low speed events are emitted by cars as the speed drops below a threshold.

The observer wants to see not the individual events but the positions and sizes of the clusters as possible indicators of traffic jams (top) or storm-affected regions (bottom).
Real-time monitoring
Visualisation of a current cluster state and its recent history

A: the current position of the cluster centre; the circle size represents the number of member events
B: the spatial convex hull of the latest state of the cluster
C: the trajectory of the cluster centre made during the observation time window
D: the spatial convex hull of all cluster states attained during the observation time window
E: cluster state details are accessible through mouse pointing
Visualisation of the cluster evolution
Our approach

1) Organize incoming events into micro-clusters

2) Join micro-clusters having \( \geq k \) connecting events

3) Treat clusters* with \( \geq N_{\text{min}} \) events as significant → present to the observer

4) Store cluster history (summarized cluster states)

* macro-clusters (= unions of micro-clusters) and isolated micro-clusters
Extensions

1) Account for thematic attributes:
   a) Add an event to a micro-cluster only if its attribute value differs from the values of the latest M events by $\leq D_{\text{max}}$.
   b) Join two micro-clusters only if their connecting events satisfy condition (a) for both micro-clusters.

   E.g., thematic attribute ‘movement direction’ for low speed car movement events.

2) Account for the number of distinct event sources:
   - The events have an attribute indicating the event source: car identifier, Twitter user identifier, ...
   - A list of distinct event sources is created and maintained for each micro and macro-cluster.
   - A cluster is treated as significant only if the number of distinct event sources is $\geq S$. 

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Start time</th>
<th>Duration</th>
<th>Movement direction (degree)</th>
<th>Cardinal direction (text)</th>
<th>N distinct event sources</th>
<th>N current events</th>
<th>N all events</th>
<th>N merged groups</th>
<th>N radius</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>0404/2007 05:17:23</td>
<td>2.30</td>
<td>E</td>
<td>106 E</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.00</td>
</tr>
<tr>
<td>19</td>
<td>0404/2007 05:19:41</td>
<td>1.25</td>
<td>E</td>
<td>106 E</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1.04</td>
</tr>
<tr>
<td>20</td>
<td>0404/2007 05:22:53</td>
<td>0.82</td>
<td>E</td>
<td>110 E</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4.32</td>
</tr>
<tr>
<td>21</td>
<td>0404/2007 05:24:49</td>
<td>0.42</td>
<td>E</td>
<td>106 E</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4.32</td>
</tr>
<tr>
<td>22</td>
<td>0404/2007 05:25:31</td>
<td>0.03</td>
<td>E</td>
<td>106 E</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4.32</td>
</tr>
<tr>
<td>23</td>
<td>0404/2007 05:25:31</td>
<td>1.63</td>
<td>E</td>
<td>110 E</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3.32</td>
</tr>
<tr>
<td>24</td>
<td>0404/2007 05:26:55</td>
<td>0.55</td>
<td>E</td>
<td>106 E</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3.15</td>
</tr>
<tr>
<td>25</td>
<td>0404/2007 05:27:23</td>
<td>0.05</td>
<td>E</td>
<td>106 E</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4.05</td>
</tr>
<tr>
<td>26</td>
<td>0404/2007 05:27:23</td>
<td>0.22</td>
<td>E</td>
<td>106 E</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4.05</td>
</tr>
<tr>
<td>27</td>
<td>0404/2007 05:28:33</td>
<td>0.13</td>
<td>E</td>
<td>106 E</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>4.15</td>
</tr>
<tr>
<td>28</td>
<td>0404/2007 05:28:45</td>
<td>0.37</td>
<td>E</td>
<td>107 E</td>
<td>8</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>255.26</td>
</tr>
<tr>
<td>29</td>
<td>0404/2007 05:30:03</td>
<td>0.48</td>
<td>E</td>
<td>101 E</td>
<td>8</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>241.17</td>
</tr>
<tr>
<td>30</td>
<td>0404/2007 05:30:40</td>
<td>0.06</td>
<td>E</td>
<td>101 E</td>
<td>8</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>241.17</td>
</tr>
<tr>
<td>31</td>
<td>0404/2007 05:30:40</td>
<td>0.47</td>
<td>E</td>
<td>101 E</td>
<td>8</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>241.17</td>
</tr>
<tr>
<td>32</td>
<td>0404/2007 05:31:15</td>
<td>0.03</td>
<td>E</td>
<td>106 E</td>
<td>8</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>215.56</td>
</tr>
<tr>
<td>33</td>
<td>0404/2007 05:31:21</td>
<td>0.00</td>
<td>E</td>
<td>106 E</td>
<td>8</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>215.56</td>
</tr>
<tr>
<td>34</td>
<td>0404/2007 05:31:21</td>
<td>0.00</td>
<td>E</td>
<td>106 E</td>
<td>8</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>215.56</td>
</tr>
<tr>
<td>35</td>
<td>0404/2007 05:31:21</td>
<td>0.00</td>
<td>E</td>
<td>106 E</td>
<td>8</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>215.56</td>
</tr>
<tr>
<td>36</td>
<td>0404/2007 05:32:57</td>
<td>0.38</td>
<td>W</td>
<td>206 W</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2.85</td>
</tr>
<tr>
<td>37</td>
<td>0404/2007 05:28:45</td>
<td>0.48</td>
<td>W</td>
<td>206 W</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2.85</td>
</tr>
<tr>
<td>38</td>
<td>0404/2007 06:30:15</td>
<td>0.02</td>
<td>W</td>
<td>271 W</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>65.72</td>
</tr>
<tr>
<td>39</td>
<td>0404/2007 06:32:23</td>
<td>0.00</td>
<td>W</td>
<td>271 W</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>63.13</td>
</tr>
<tr>
<td>40</td>
<td>0404/2007 06:32:23</td>
<td>0.00</td>
<td>W</td>
<td>271 W</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>63.13</td>
</tr>
</tbody>
</table>

[Diagram showing data points with labels and colors, possibly indicating clusters or events]
Example: simulated real time clustering of storm-related tweet events
Example: simulated real time clustering of storm-related tweet events
Resulting clusters
Cluster trajectories
Comparison to “classical” density-based clustering (OPTICS; not accounting for N of distinct sources)
Comparison to “classical” density-based clustering (OPTICS; not accounting for N of distinct sources)
Summary

• Types of spatio-temporal data: spatial events, trajectories (quasi-continuous, episodic), spatial time series

• Two aspects of spatial time series: local time series referring to places vs. spatial situations referring to times

• Transformations of spatio-temporal data

• Partition-based vs. density-based clustering

• Two-way partition-based clustering of spatial time series:
  • clustering of places by similarity of local time series
  • clustering of times by similarity of spatial situations

• Density-based clustering of spatial events for detecting spatio-temporal concentrations

• Density-based clustering of streaming events