Multidimensional and Multilevel Analysis of International Media Flows

Robin Lamarche-Perrin, Claude Grasland

Collaborators from ANR GEOMEDIA:
Yves Demazeau
Benjamin Loveluck
Hugues Pecout
Jean-Marc Vincent
What is the state of the world?

International sections of daily newspapers

Which event made the news?
- At a given date?
- Regarding a given country?
- According to a given newspaper?

⇒ Geomedia agenda-setting

ANR GEOMEDIA Project: Data and tools for a quantitative analysis of the multiple facets of international news at a worldwide scale
→ Media studies, political sciences, quantitative geography, computer science
The Three Dimensions of Media Flows

International Media Flows

THE GUARDIAN

THE TIMES OF INDIA

Paper 1

Paper 2

Paper 3

“Japan”

“Madrid”

“French”

“Spain”

Weighted temporal bipartite graph

Example:
1 article published by media $m_1$ at time $t_1$ and citing country $s_1$
1 article published by media $m_4$ at time $t_1$ and citing country $s_4$
1 article published by media $m_4$ at time $t_1$ and citing countries $s_3$ and $s_4$
The Three Dimensions of Media Flows

Geomedical Cube
(media × space × time)

Pays (S) → Cuba
Le Monde
New York Times
China Daily

Temps (T) → Semaine 1
Semaine 2
Semaine 3

 Média (M) → m1
m2
m3
m4

Weighted temporal bipartite graph

Time

Data: 292,767 articles
published by 36 newspapers (in 23 different states)
during 52 weeks (from 28/04/2014 to 26/04/2015)
and citing 197 countries (recognised by the UN)
What can be said about one edge?

Cuba as been cited **16 times** by Le Monde during the week of December 15th, 2014.

→ **Is that a lot?**

Knowing that Le Monde made 276 citations **that week**, and that it usually devotes 0.55% of its citations to Cuba?

Knowing that Le Monde made 276 citations **that week**, and that 9.0% of citations of all media **that week** were dedicated to Cuba?

Knowing that Le Monde made 64 citations about Cuba within the whole corpus, and that 18.8% of all citations about Cuba where concentrated on **that particular week**?
Part I

The Multiple Facets of Media Flows

Knowing that Le Monde made 276 citations that week, and that it usually devotes 0.55% of its citations to Cuba?

→ One would then expect 1.5 citations, so 16 is a lot!

Given a media \( m \in M \), detect spatio-temporal irregularities \((s_j, t_k) \in S \times T\).

\[
\begin{array}{|c|c|c|c|}
\hline
m & s_1 & s_2 & s_3 \\
\hline
m_1 & 3 & 4 & 3 \\
\hline
m_2 & 3 & 5 & 4 \\
\hline
m_3 & 2 & 7 & 1 \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|c|}
\hline
m & s_1 & s_2 & s_3 \\
\hline
m_1 & & & 10 \\
\hline
m_2 & & & 12 \\
\hline
m_3 & & & 10 \\
\hline
\end{array}
\]
Knowing that Le Monde made 276 citations that week, and that it usually devotes 0.55% of its citations to Cuba?

→ One would then expect 1.5 citations, so 16 is a lot!

Given a media $m \in M$, detect spatio-temporal irregularities $(s_j, t_k) \in S \times T$.

\[
\begin{array}{c|ccc}
 m & s_1 & s_2 & s_3 \\
 \hline
 t_1 & 3 & 4 & 3 \\
 t_2 & 3 & 5 & 4 \\
 t_3 & 2 & 7 & 1 \\
\end{array}
\]

\[
\begin{array}{c|ccc}
 m & s_1 & s_2 & s_3 \\
 \hline
 t_1 & +.51 & -.12 & +.51 \\
 t_2 & 0 & -.11 & +.63 \\
 t_3 & -.09 & +.73 & -.43 \\
\end{array}
\]

Raw values: $v(m, s_j, t_k)$

Significativity: $p(m, s_j, t_k) = \Pr(X \geq v(m, s_j, t_k))$

with $X \sim \text{Pois}(v^*(m, s_j, t_k))$

\[
\sigma(m, s_j, t_k) = 2p(m, s_j, t_k) - 1 \in [-1, +1]
\]

Expected values:

\[
v^*(m, s_j, t_k) = \frac{v(m, s_j, \cdot)}{v(m, \cdot, \cdot)}
\]
This data model measures significant divergences with respect to the **mean internal agenda** of a particular newspaper. 

*(Le Monde, week of 15th December, 2014)*
Knowing that Le Monde made 276 citations that week, and that 9.0% of citations of all media that week were dedicated to Cuba?

→ One would then expect 25 citations, so 16 is actually not much...

Given a time $t \in \mathcal{t}$, detect spatio-media irregularities $(m_i, s_j) \in M \times S$. 

<table>
<thead>
<tr>
<th></th>
<th>$s_1$</th>
<th>$s_2$</th>
<th>$s_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_1$</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>$m_2$</td>
<td>3</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>$m_3$</td>
<td>2</td>
<td>7</td>
<td>1</td>
</tr>
</tbody>
</table>

Raw values: $v(m_i, s_j, t)$

Marginal values:

$\displaystyle v(m_i, \ldots, t) = \sum_{m_i \in M} v(m_i, s_j, t)$

$\displaystyle v(\ldots, s_j, t) = \sum_{s_j \in S} v(m_i, s_j, t)$

$\displaystyle v(\ldots, \ldots, t) = \sum_{m_i \in M} v(m_i, s_j, t)$

Expected values: $v^*(m_i, s_j, t) = \frac{v(m_i, \ldots, t) v(\ldots, s_j, t)}{v(\ldots, \ldots, t)}$
This data model measures significant divergences with respect to the **mean spatial agenda** of all media on a particular time period.

*(Le Monde, week of December 15th, 2014)*
Knowing that Le Monde made 64 citations about Cuba within the whole corpus, and that 18.8% of all citations about Cuba were concentrated on that particular week?

→ One would then expect 12 citations, so 16 is a little bit more.

Given a country \( s \in S \), detect tempo-media irregularities \((m_i, t_k) \in M \times T\).

<table>
<thead>
<tr>
<th>( s )</th>
<th>( t_1 )</th>
<th>( t_2 )</th>
<th>( t_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_1 )</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>( m_2 )</td>
<td>3</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>( m_3 )</td>
<td>2</td>
<td>7</td>
<td>1</td>
</tr>
</tbody>
</table>

Raw values:

\[ v(m_i, s, t_k) \]

Marginal values:

\[ v(m_i, s, \cdot) = \sum_{m_i \in M} v(m_i, s, t_k) \]
\[ v(\cdot, s, t_k) = \sum_{t_k \in T} v(m_i, s, t_k) \]
\[ v(\cdot, \cdot) = \sum_{m_i \in M} v(m_i, s, t_k) \]

Expected values:

\[ v^*(m_i, s, t_k) = \frac{v(m_i, s, \cdot) v(\cdot, s, t_k)}{v(\cdot, \cdot)} \]
This data model measures significant divergences with respect to the **mean temporal agenda** of all media regarding a particular country (Ukraine, Le Monde)
Part II

The Multiple Scales of Media Flows

Looking for Geographical Scales

ISTA Model

Le Monde

July 2011

MICROSCOPIC REPRESENTATION

Aggregation

Disaggregation

R. Lamarche-Perrin, C. Grasland

Multidimensional and Multilevel Analysis of Media Flows
Looking for Geographical Scales

ISTA Model
Le Monde
July 2011

International Event
Detection of Media Events
MICROSCOPIC REPRESENTATION
Newspaper LE MONDE in July 2011

<table>
<thead>
<tr>
<th>m</th>
<th>s₁</th>
<th>s₂</th>
<th>s₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>t₁</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>t₂</td>
<td>3</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>t₃</td>
<td>2</td>
<td>7</td>
<td>1</td>
</tr>
</tbody>
</table>

Aggregation

<table>
<thead>
<tr>
<th>m</th>
<th>s₁</th>
<th>s₂</th>
<th>s₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>t₁</td>
<td>3</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>t₂</td>
<td>3</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>t₃</td>
<td>2</td>
<td>8</td>
<td>10</td>
</tr>
</tbody>
</table>

R. Lamarche-Perrin, C. Grasland
Multidimensional and Multilevel Analysis of Media Flows
Looking for Geographical Scales

ISTA Model
Le Monde
July 2011

MESOSCOPIC REPRESENTATION

Aggregation

R. Lamarche-Perrin, C. Grasland
Multidimensional and Multilevel Analysis of Media Flows
Looking for Geographical Scales

ISTA Model
Le Monde
July 2011

MACROSCOPIC REPRESENTATION

P2: The Levels of Representation
Newspaper LE MONDE in July 2011

Disaggregation →

Aggregation →

R. Lamarche-Perrin, C. Grasland
Multidimensional and Multilevel Analysis of Media Flows
Looking for Geographical Scales

ISTA Model
Le Monde
July 2011

MULTIRESOLUTION REPRESENTATION

P2: The Levels of Representation
Newspaper LE MONDE in July 2011

Aggregation
Disaggregation

National Event
International Event

R. Lamarche-Perrin, C. Grasland
Multidimensional and Multilevel Analysis of Media Flows
Data Aggregation and Information Loss

**Initial Data**

\[ X \in \{ x_1, \ldots, x_n \} \]

\[ p(x_i) = \Pr(X = x_i) \]

**Information Loss**

\[ D_{KL}(p\|q) = \sum_i p(x_i) \log_2 \frac{p(x_i)}{q(x_i)} \]

**Aggregation**

\[ \hat{X} \in \{ \hat{x}_1, \ldots, \hat{x}_m \} \]

\[ p(\hat{x}_j|x_i) = \Pr(\hat{X} = \hat{x}_j|X = x_i) \]

**Disaggregation**

\[ X^* \in \{ x_1, \ldots, x_n \} \]

\[ u(x_i) = \Pr(X^* = x_i) \]

\[ q(x_i|\hat{x}_j) = \frac{p(\hat{x}_j|x_i)u(x_i)}{\sum_{i'} p(\hat{x}_{i'}|x_{i'})u(x_{i'})} \]

\[ q(x_i) = \sum_j q(x_i|\hat{x}_j)p(\hat{x}_j) \]

**Aggregated Data**

\[ p(\hat{x}_j) = \sum_i p(\hat{x}_j|x_i)p(x_i) \]
Multidimensional Aggregation in the Cube

No privileged dimension

One privileged dimension

Two privileged dimensions

R. Lamarche-Perrin, C. Grasland
Multidimensional and Multilevel Analysis of Media Flows
Media Aggregation

countries

MSA Model
Week of 15/12/2014

R. Lamarche-Perrin, C. Grasland
Multidimensional and Multilevel Analysis of Media Flows
### Media × Time Aggregation

**No information loss → 144 aggregates**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<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>en_IND_tindia_int</td>
<td>120</td>
<td>119</td>
<td>118</td>
<td>117</td>
<td>116</td>
<td>115</td>
<td>114</td>
<td>113</td>
<td>112</td>
<td>111</td>
<td>110</td>
<td>109</td>
<td></td>
</tr>
<tr>
<td>en_USA_nytime_int</td>
<td>60</td>
<td>59</td>
<td>58</td>
<td>57</td>
<td>56</td>
<td>55</td>
<td>54</td>
<td>53</td>
<td>52</td>
<td>51</td>
<td>50</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>en_AUS_austria_int</td>
<td>72</td>
<td>71</td>
<td>70</td>
<td>69</td>
<td>68</td>
<td>67</td>
<td>66</td>
<td>65</td>
<td>64</td>
<td>63</td>
<td>62</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>en_CHN_mopost_int</td>
<td>108</td>
<td>107</td>
<td>106</td>
<td>105</td>
<td>104</td>
<td>103</td>
<td>102</td>
<td>101</td>
<td>100</td>
<td>99</td>
<td>98</td>
<td>97</td>
<td></td>
</tr>
<tr>
<td>en_JPN_jatime_int</td>
<td>132</td>
<td>131</td>
<td>130</td>
<td>129</td>
<td>128</td>
<td>127</td>
<td>126</td>
<td>125</td>
<td>124</td>
<td>123</td>
<td>122</td>
<td>121</td>
<td></td>
</tr>
<tr>
<td>en_NZL_nzhera_int</td>
<td>96</td>
<td>95</td>
<td>94</td>
<td>93</td>
<td>92</td>
<td>91</td>
<td>90</td>
<td>89</td>
<td>88</td>
<td>87</td>
<td>86</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>en_GBR_guardian_int</td>
<td>48</td>
<td>47</td>
<td>46</td>
<td>45</td>
<td>44</td>
<td>43</td>
<td>42</td>
<td>41</td>
<td>40</td>
<td>39</td>
<td>38</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>en_SGP_twoday_int</td>
<td>144</td>
<td>143</td>
<td>142</td>
<td>141</td>
<td>140</td>
<td>139</td>
<td>138</td>
<td>137</td>
<td>136</td>
<td>135</td>
<td>134</td>
<td>133</td>
<td></td>
</tr>
<tr>
<td>en_CAN_starca_int</td>
<td>84</td>
<td>83</td>
<td>82</td>
<td>81</td>
<td>80</td>
<td>79</td>
<td>78</td>
<td>77</td>
<td>76</td>
<td>75</td>
<td>74</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td>fr_FRA_lémont_int</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>es_ESP_elpais_int</td>
<td>36</td>
<td>35</td>
<td>34</td>
<td>33</td>
<td>32</td>
<td>31</td>
<td>30</td>
<td>29</td>
<td>28</td>
<td>27</td>
<td>26</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>fr_BEL_lesoir_int</td>
<td>24</td>
<td>23</td>
<td>22</td>
<td>21</td>
<td>20</td>
<td>19</td>
<td>18</td>
<td>17</td>
<td>16</td>
<td>15</td>
<td>14</td>
<td>13</td>
<td></td>
</tr>
</tbody>
</table>

**MSA Model**

Ukraine
0.7% of information loss → 50 aggregates

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<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>en_IND_tindia_int</td>
<td>49</td>
<td>44</td>
<td>39</td>
<td>36</td>
<td>29</td>
<td>25</td>
<td>19</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>en_USA_nytime_int</td>
<td>47</td>
<td>47</td>
<td>38</td>
<td>36</td>
<td>36</td>
<td>23</td>
<td>19</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>en_AUS_austria_int</td>
<td>48</td>
<td>42</td>
<td>39</td>
<td>36</td>
<td>23</td>
<td>19</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>en_CHN_mopost_int</td>
<td>49</td>
<td>46</td>
<td>38</td>
<td>36</td>
<td>23</td>
<td>19</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>en_JPN_jatime_int</td>
<td>50</td>
<td>46</td>
<td>43</td>
<td>38</td>
<td>23</td>
<td>19</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>en_NZL_nzhera_int</td>
<td>50</td>
<td>46</td>
<td>43</td>
<td>38</td>
<td>23</td>
<td>19</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>en_GBR_guardi_int</td>
<td>50</td>
<td>46</td>
<td>43</td>
<td>38</td>
<td>23</td>
<td>19</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>en_SGP_twoday_int</td>
<td>50</td>
<td>46</td>
<td>43</td>
<td>38</td>
<td>23</td>
<td>19</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>en_CAN_starca_int</td>
<td>50</td>
<td>46</td>
<td>43</td>
<td>38</td>
<td>23</td>
<td>19</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>fr_FRA_lemonde_int</td>
<td>50</td>
<td>46</td>
<td>43</td>
<td>38</td>
<td>23</td>
<td>19</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>es_ESP_elpais_int</td>
<td>50</td>
<td>46</td>
<td>43</td>
<td>38</td>
<td>23</td>
<td>19</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>fr_BEL_lesoir_int</td>
<td>50</td>
<td>46</td>
<td>43</td>
<td>38</td>
<td>23</td>
<td>19</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>4</td>
</tr>
</tbody>
</table>

Signif. (0-400)
2.9% of information loss → 29 aggregates
### Ukraine

- 5.9% of information loss → 16 aggregates

#### Media × Time Aggregation

<table>
<thead>
<tr>
<th>Media</th>
<th>Time</th>
<th>Signif.</th>
</tr>
</thead>
<tbody>
<tr>
<td>en_IND_tindia_int</td>
<td>2013-12-30</td>
<td>-400</td>
</tr>
<tr>
<td>en_USA_nytime_int</td>
<td>2014-01-30</td>
<td>0</td>
</tr>
<tr>
<td>en_AUS_austria_int</td>
<td>2014-02-27</td>
<td>200</td>
</tr>
<tr>
<td>en_CHN_mopost_int</td>
<td>2014-03-30</td>
<td>-400</td>
</tr>
<tr>
<td>en_JPN_jatime_int</td>
<td>2014-04-30</td>
<td>0</td>
</tr>
<tr>
<td>en_NZL_nzhera_int</td>
<td>2014-05-30</td>
<td>0</td>
</tr>
<tr>
<td>en_GBR_guardian_int</td>
<td>2014-06-29</td>
<td>0</td>
</tr>
<tr>
<td>en_SGP_twoday_int</td>
<td>2014-07-30</td>
<td>0</td>
</tr>
<tr>
<td>en_CAN_starca_int</td>
<td>2014-08-29</td>
<td>0</td>
</tr>
<tr>
<td>fr_FRA_lmonde_int</td>
<td>2014-09-29</td>
<td>0</td>
</tr>
<tr>
<td>es_ESP_elpais_int</td>
<td>2014-10-30</td>
<td>0</td>
</tr>
<tr>
<td>frBEL_lesoir_int</td>
<td>2014-11-29</td>
<td>0</td>
</tr>
</tbody>
</table>
8.5% of information loss → 11 aggregates

Multidimensional and Multilevel Analysis of Media Flows
18.3% of information loss → 5 aggregates

MSA Model

Ukraine
Conclusion

**Geomedia Cube**

Events are defined as outliers with respect to the selected corpus.

Aggregation should also take into account the semantics of the three dimensions (rdv tomorrow at 9:50 in Session 5).

Measure of statistical significance could be improved (now using Poisson counting processes).

We need heuristics (sub-optimal algorithms) to work on the whole cube.

Limitations and Perspectives

**Ressources**

- Aggregation algorithms: [github.com/Lamarche-Perrin/optimal.partition](https://github.com/Lamarche-Perrin/optimal.partition)
- GEOMEDIA Project: [geomedia.hypotheses.org](http://geomedia.hypotheses.org)
- GEOMEDIA Data: [www.gis-cist.fr/en/mise-a-disposition-dun-echantillon-de-la-base-de-donnees-geomedia](http://www.gis-cist.fr/en/mise-a-disposition-dun-echantillon-de-la-base-de-donnees-geomedia)
« Aujourd’hui, nous sommes confrontés à un autre infini : l’infiniment complexe. Mais cette fois, plus d’instrument. »