

WIDE Technical-Report in 2006

トラフィックローカリティの可  
視化-試行錯誤の過程-  
wide-tr-netviz-trafficviz-02.pdf



WIDE Project : <http://www.wide.ad.jp/>

*If you have any comments on this document, please contact to [ad@wide.ad.jp](mailto:ad@wide.ad.jp)*

トラフィックローカリティの可視化

－試行錯誤の過程－

Wide-tr-tafficviz-02.pdf

### 0.背景

- 1.可視化とは
- 2.最適化までの試行錯誤 – サンプルデータ
- 3.最適化までの試行錯誤 – 正式データ
- 4.トラフィックローカリティの可視化 – 完成へ
- 5.まとめ

### 0.背景

可視化とは、見えないものを見えるようにすることである。しかし、そこから言いたいことを伝達するためには、どのような見せ方があるか、どのような手法がよいのかを検討すべき、との考えから「国内インターネットのトラフィックマトリクス」をテーマにトラフィックローカリティの可視化を試みた。

6月の netviz ワークショップ、9月の WIDE 合宿などでの議論を含め、サンプルデータのみがあるという状態から最終的な形に整えるまでの試行錯誤の過程を報告する。

### 1.可視化とは

何を見せるか＝何を言いたいか。

差、同様、特異などいろいろある。これは情報を可視化する上で最も重要なポイントであり、すなわち可視化とは「表現を最適化する」ということだと考える。表現を最適化するためには、いくつかの段階を踏み、必要に応じて戻りながら進めていくべきである。

- ◇ Phase1：何が見えるか
- ◇ Phase2：何を見せるか
- ◇ Phase3：どんな手法にするか  
(グラフ、3DCG、アニメーションなど)

- ◇ Phase4：どのように見せるか  
(色、質感、テイストなど)

また、表現の傾向として大きく2つに分けることができる。

#### ○具体的な傾向

正確さを重視し、わかりやすく簡潔にまとめる。(数字の羅列、目盛など)

#### ○抽象的な傾向

見え方を重視し、グラフィカルな要素を強くする。(格好よく、きれいになど)

今回は後者の抽象的な傾向に注力して進める。

## 2.最適化までの試行錯誤 – サンプルデータ

まずは、phase1 どのように見えるのか。各都道府県の総トラフィック量のサンプルデータ（図 2.1）を元にボリュームをつかむ。図 2.1 の通りに北海道から順に上から並べ、IN と OUT をそれぞれ左右に配置し棒グラフを作成した。（図 2.2）

〈要素〉

- ・ 47 都道府県
- ・ IN：各県に入ってくる総トラフィック量
- ・ OUT：各県から出ていく総トラフィック量

			IN	OUT
1	北海道	hokkaido	833.3133541	703.328458
2	青森県	aomori	118.5739343	97.39717566
3	岩手県	iwate	137.1537408	94.25967554
4	宮城県	miyagi	278.709743	228.8963367
5	秋田県	akita	119.7014571	104.6289502
6	山形県	yamagata	93.64432145	77.75200704
7	福島県	fukushima	211.3554335	158.3476694
8	茨城県	ibaraki	167.7697585	159.3878722
9	栃木県	tochigi	178.882625	174.3752537
10	群馬県	gunma	221.9218553	191.0498037
11	埼玉県	saitama	868.223687	914.8335375
12	千葉県	chiba	732.2393839	722.7037926
13	東京都	tokyo	2025.111421	1950.629774
14	神奈川県	kanagawa	1153.189676	1286.570861

図 2.1

東京都の数値が圧倒的に大きいことがわかる。その他、神奈川県、愛知県などの主要都市の数値が高い。現時点では目盛や単位がないため、各県の微妙な差がわかる程度で、数値を把握することができない。

つぎに、「見せたいものが何か」を探るために、どんな見せ方があるか、いくつか作成した。（Phase2.3.4 の反復）A から E の 5 種類を作成し、このうちダッシュがついているものは、議論の内容を反映させて更新している。

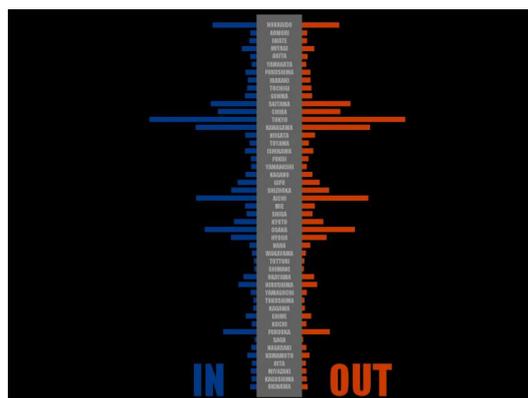


図 2.2

### 【タイプ A ・ シンボリックなイメージ】 図 2.3

地理情報として日本地図を使用することによってシンボリックなイメージになり、感覚で地理情報を把握できる。IN（入ってくる）と OUT（出ていく）をイメージできるように、それぞれ IN を下方に、OUT を上方に高さを与え、3次元で表現。

〈要素〉

- ・ 各県の総トラフィック量（IN/OUT）
- ・ 都道府県（日本地図を使用）

〈評価〉

視点を固定させると見えない部分が出てくる。

また、回転させると見慣れない地図の景色になるため、わかりづらい。

IN と OUT については、天地をつけるとマイナスのイメージにつながり、嫌がられる場合がある。

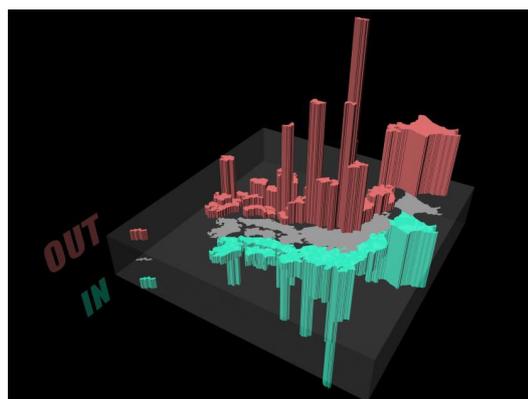


図 2.3

### 【タイプ B】 図 2.4

軸を 1 つ加えることによって興味深いデータが生まれるのではないかと思います、一人当たりのトラフィック量を各県の人口から算出し、都道府県、各県の総トラフィック量に加えて 3 つ目の軸として 3次元で表現してみる。3つの軸が交わる場所に IN と OUT のポイントをそれぞれ打った。

〈要素〉

- ・ X : 都道府県
- ・ Y : 各県 1 人当たりのトラフィック量
- ・ Z : 各県の総トラフィック量
- ・ ポイント : IN/緑、OUT/赤

〈評価〉

単位や目盛がないため、何を示しているかがわからない。また、ポイントが重なって見えにくく、各県の比較がしにくい。

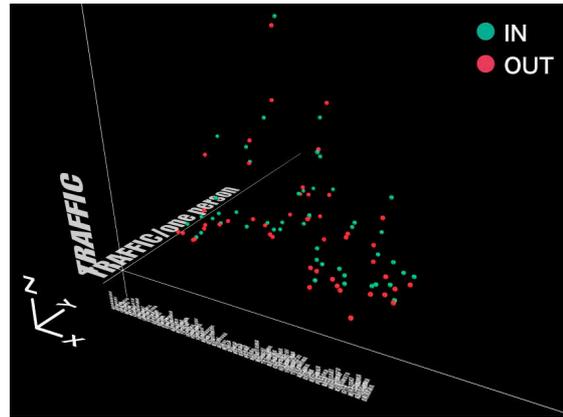


図 2.4

### 【タイプ B'】 図 2.5

〈改善点〉

- ・ 各県の数値の差をわかりやすくするために、スケールを 4 倍にする
- ・ 軸の方向を変えて読み取りやすくする
- ・ 背景を明るくし、暗い印象を払拭

〈要素〉

- ・ X : 各県の総トラフィック量
- ・ Y : 各県 1 人当たりのトラフィック量
- ・ Z : ソース・都道府県
- ・ ポイント : IN/紫、OUT/緑

〈評価〉

スケールを調整してもまだポイントが重なって効果がない。スケールの調整よりもそれぞれのサイズや形状を変える方が効果が出るのではないか。静止画としてみるには、都道府県が細かすぎる。また、新たに軸を加えると趣旨が混同するため、加える必要はない。

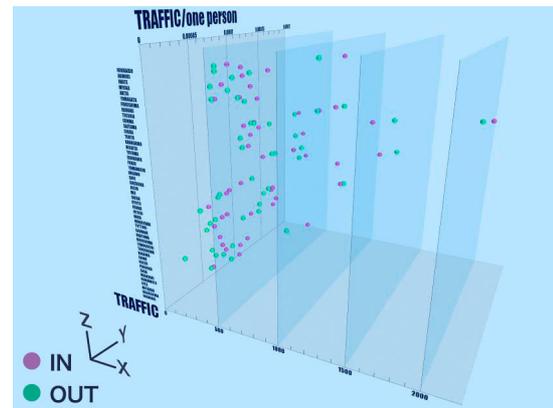


図 2.5

### 【タイプ C】 図 2.6

たくさんの要素を一度に見せたらどうなるか試してみる。図 2.1 を元に各県の総トラフィック量に比例して配分されるとして県間トラフィックデータを算出し、47 都道府県では細かすぎるため、東北・関東・北陸・中部・近畿・中国・四国・九州の 8 エリアに区分した。ソースを X、デスティネーションを Y に取ってマトリクスに示し、各エリアに配分される値を Z として頂点に面を張った。効果的なライティングを施せば、僅かな値の差や全体のボリュームがわかりやすい。

〈要素〉

- ・ 都道府県を 8 エリアに区分
- ・ 3 つの軸
  - X : ソース (8 エリア)
  - Y : デスティネーション (8 エリア)
  - Z : トラフィック量
- ・ 棒グラフ : 各エリアのトラフィック (IN/OUT)
- ・ ポイント : 各エリアに閉じているトラフィック

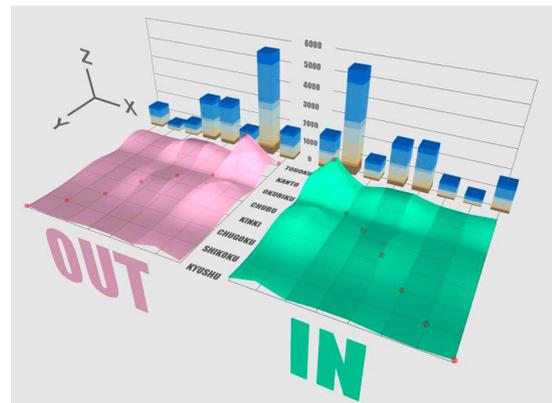


図 2.6

〈評価〉

見る方向によってさまざまな情報を得ることができることを狙ったが、要素を盛り込みすぎてどこを見ればよいかわからなくなり、「何を見せたいのか」が明確でない悪い例。

【タイプ D】 図 2.7

モザイク画のようなイメージで、行 (OUT) と列 (IN) の幅を等しくしたマトリクスを着色し、彩度と明度で表現した。

〈要素〉

- ・各県の総トラフィック量 IN (横軸)
- ・各県の総トラフィック量 OUT (縦軸)

〈評価〉

微妙な変化がわかって興味深いのが、0~90、100~1000 の2つに分けて配色したために割り当て方が不適切で誤解を招く恐れがある。0~1000 の間でグラデーションの配色をするべきである。

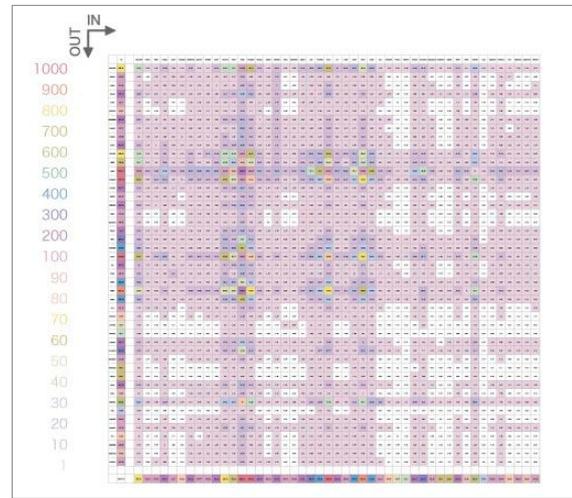


図 2.7

【タイプ E】 図 2.8

これまで IN と OUT のトラフィック総量が異なるためにそれぞれ作図していたが、量ではなく地域間の分布の割合に的を絞り、要素をまとめてみる。8 エリア間のトラフィック分布を縦軸にソース、横軸にデスティネーションを取り、通信相手の分布を行と列で示す。ソース、デスティネーションそれぞれの各エリアに正規化した割合を線の太さ、トラフィック量を色で表現した。微妙な変化が見えるのではないかと期待した。

〈要素〉

- ・都道府県 (8 エリア)
- ・トラフィック量 (色)
- ・ソース (縦軸)
- ・デスティネーション (横軸)
- ・ソース/デスティネーションの各エリアに正規化した割合 (線の太さ)

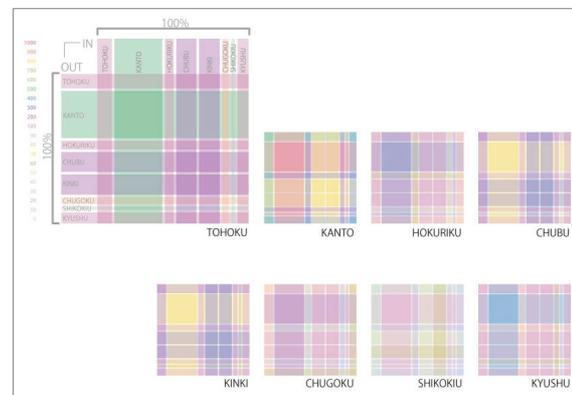


図 2.8

〈評価〉

各エリアによって柄がちがっておもしろいので、47 都道府県をやってみてはどうだろうか。チェック柄を意識しすぎたため、色を半透明にすると縦軸と横軸が重なった部分の色に対して説明がつかない。

●ここまで

○各県の総トラフィック量を正規化

各県のトラフィック量の値の差が大きすぎるため、総トラフィック量 (IN/OUT) を一度に見せるのは難しい。全体を表示している状態での最大値と最小値の比較は困難であるため、合計値を正規化してスケールを均一にし、割合を見る方がよいのではないかと。

○47 都道府県を 8 エリアに

47 都道府県を並べただけではわかりづらい。日本地図は都道府県を表記する必要はなくなるが、配置換えができず、バリエーションには限りがある。8 エリア (東北・関東・北陸・中部・近畿・中国・四国・九州) にわけた方がよいのではないかと。

○色への配慮

背景を黒にすると、コントラストがはっきりするが暗い印象になる。一方、明るい配色のみで構成すると区別がつきにくい。

○何をjせるか

- ・各県に閉じているトラフィック量
- ・「各県の地域性がない」ということ

○成果物

ポスターを作成する。

3.最適化までの試行錯誤 – 正式データ

2005 年 7 月 4 日から 7 月 10 日までの一週間、ある国内 ISP の全国ブロードバンドユーザを対象に、アクセス網を収容しているルータ群で、Sampled NetFlow 機能を用いて収集されたデータが届いた時点でこれを正式データとし、サンプルデータで作成したものをベースに差し替えて進めた。前の項目で述べたとおり、何をjせたいか、どう表現するか要素がまとまってきたが、Phase2.3.4 を反復しながら最終形へ整えていく。

〈正式データ〉

図 3.1 に例を挙げた。ソースを縦軸、デスティネーションを横軸に取り、トラフィック量全体を 100 として各県に分布する割合を示す。A-D の部分に 47 都道府県もしくは 8 エリアが入り。srcB と dstB が交わる“18”が B に閉じているトラフィックの割合である。

- ・データ 1 – 全体に対する割合
- ・データ 2 – ソースの各県に正規化した割合
- ・データ 3 – デスティネーションの各県に正規化した割合

dst src	ALL	A	B	C	D
ALL	100	20	40	20	20
A	20	4	8	4	4
B	50	10	18	12	10
C	10	2	4	2	2
D	20	4	10	2	4

データ 1

dst src	ALL	A	B	C	D
A	100	20	40	20	20
B	100	20	36	24	20
C	100	20	40	20	20
D	100	20	50	10	20

データ 2

dst src		A	B	C	D
ALL		100	100	100	100
A		20	20	20	20
B		50	45	60	50
C		10	10	10	10
D		20	25	10	20

データ 3

図 3.1

【タイプ E'】 図 3.2

47 都道府県におけるデータ 2 を横軸 (左上)、データ 3 を縦軸 (上中央) に取り、それぞれの数値を太さで表現し、これらを重ねて、ひとつの図にまとめた。

〈改善点〉

- ・説明がつくように、色を 2 色に限定
- 黒・各県のトラフィック
- 赤・閉じているトラフィック

〈評価〉

2 色にしたことで何を見せたいかが明確になった。各県で試したが、細かすぎて変化がよくわからない。

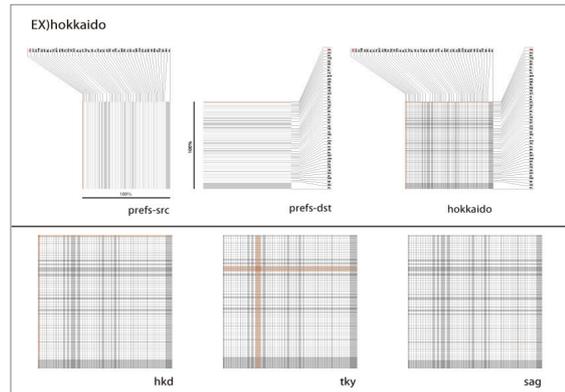


図 3.2

【タイプ F】 図 3.3

図右のように、まずは原点を中央にして左右にそれぞれ 47 都道府県におけるデータ 2 とデータ 3 を棒グラフで表した。各県に閉じたトラフィックは赤で示している。棒グラフを立体にして日本地図の県庁所在地の位置に配置した。日本地図は棒グラフの原点を意味する。

〈要素〉

- ・都道府県 (日本地図を使用)
- ・ローカリティ: 各県に閉じたトラフィック (赤)
- ・ソース: src/左・奥
- ・デスティネーション: dst/右・手前

〈評価〉

閉じているトラフィックの割合が少ないため、この部分のみを着色しても変化がわからず効果はない。重なった部分が見えない。

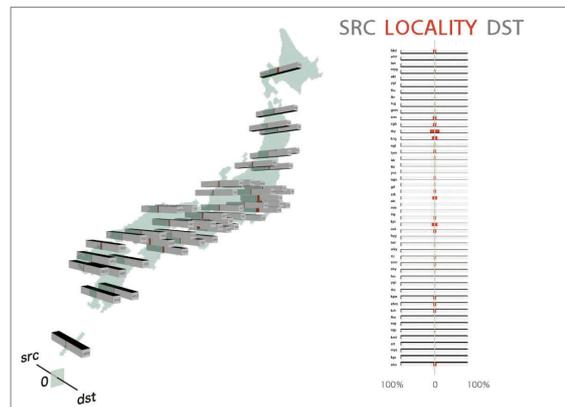


図 3.3

【タイプ G】 図 3.4

各エリアの情報を日本地図に配置。上面、正面、側面の 3 方向にそれぞれ異なる情報を表している。見る方向によって違う情報を得られる。

〈要素〉

- ・都道府県 (8 エリア)
- ・3 方向
  - X: ソースの各エリアに正規化した割合
  - Y: デスティネーションの各エリアに正規化した割合
  - Z: ローカリティ (各エリアに閉じたトラフィック)

〈評価〉

X 軸、Y 軸から見ると、ボールが重なって見えない。上面に視点を固定してボール自体を見たい方向に回転させてはどうか。

ひと目でわかるような見せ方を考えるべきだ。

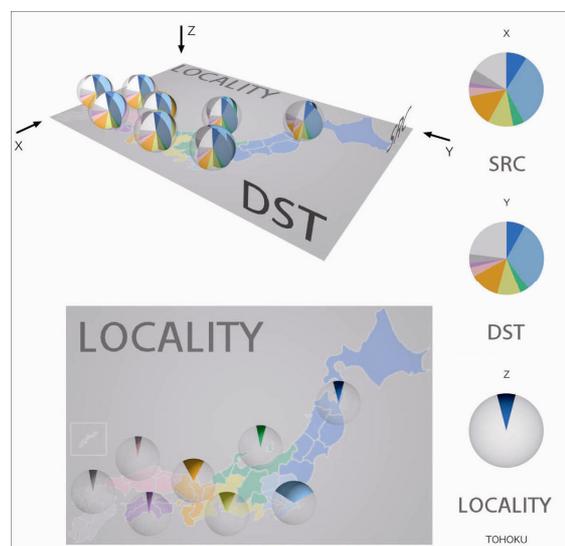


図 3.4

【タイプ H】 図 3.5

まず 2 次元で、データ 2 と 3 の数値をそれぞれ横線と縦線の折れ線で表し、エリア内に閉じたトラフィックを赤線で表現 (図 3.5 左、東北を例)。この要領で各エリアの折れ線をつくり、それぞれ日本地図の各エリアの中央に配置し (図 3.5 右)、各エリアに閉じたトラフィックをわかりやすくするために折れ線に高さを与えた (高さは意味をもたない)。

〈要素〉

- ・都道府県 (8 エリア)
- ・ソースの各エリアに正規化した割合 (縦軸)
- ・デスティネーションの各エリアに正規化した割合 (横軸)
- ・エリア内に閉じているトラフィック (赤線)

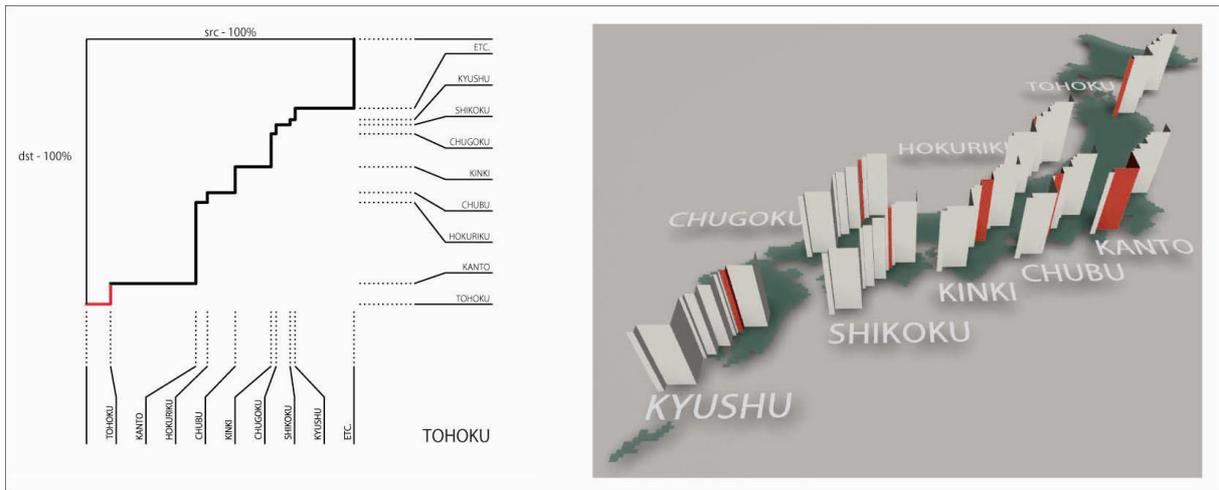


図 3.5

〈評価〉

要素が整理されていてわかりやすいが、高さに定義がないのであれば 3 次元で表現する必要はないのではないかと。3 次元ではなく、すべてをまとめて 2 次元で表したらどうか。また、上が九州ではなく東北になる方が見やすい。

【タイプ H'】 図 3.6

〈改善点〉

- ・各エリアの折れ線ひとつの図にまとめて 2 次元にし、エリアごとに色分けする
- ・各エリアに閉じたトラフィックを矩形の領域で表現

〈要素〉

- ・都道府県 (8 エリア)
- ・ソースの各エリアに正規化した割合 (縦軸)
- ・デスティネーションの各エリアに正規化した割合 (横軸)
- ・エリア内に閉じているトラフィック (矩形の領域)

〈評価〉

上が九州、下が東北なのはわかりづらいので、東北が上にくるように反転させた方がよいのではないかと。また、どこかの県を取り出して、エリアではなく各県の分布を表したらどうか。

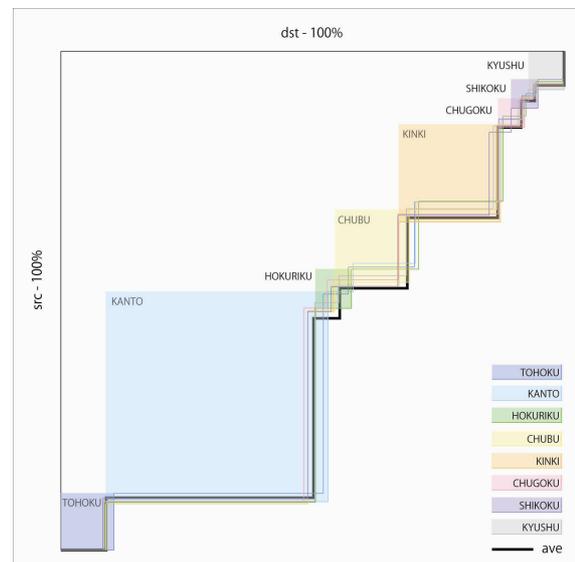


図 3.6

### 【タイプ I】 図 3.7

各県の分布の割合に図の左にデータ 2 (src 縦軸/dst 横軸)、右にデータ 3 (src 横軸/dst 縦軸) の数値を直線で表し、各県ごとに線で結んだ。

〈要素〉

- ・ 都道府県 (上から順に北海道→沖縄)
- ・ 8 エリアを色分け
- ・ ソースの各県に正規化した割合
- ・ デスティネーションの各県に正規化した割合

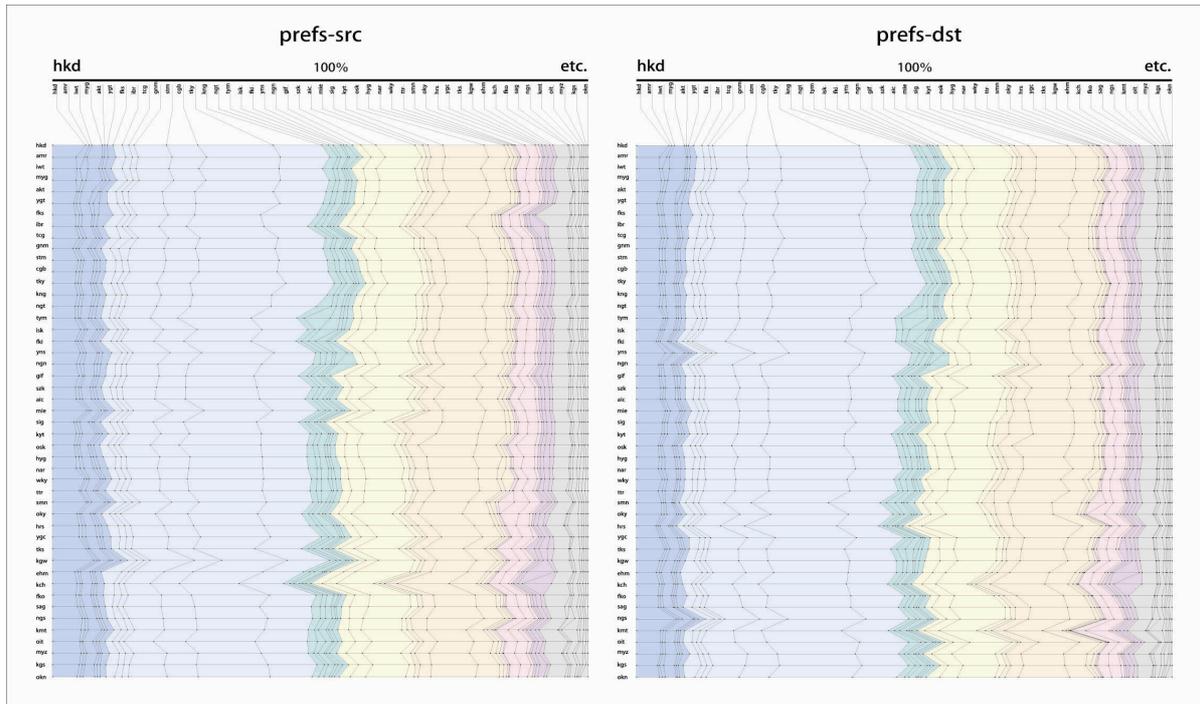


図 3.7

〈評価〉

微妙な変化がよく見えてわかりやすいが、色が淡すぎるのではないかと。

#### ●ここまで

##### ○都道府県の表現は8エリアに区分

日本地図を使用して3次元で表現することは、重なる部分が見えない場合が多く、適切ではない。3次元で作成したものを2次元に収めるのはかなり困難である。

また、47都道府県を並べると要素が多くなるため、8エリアに区分した方がわかりやすい。

##### ○何をさせるか

- ・ 各エリアとも地域性がないこと
- ・ 各エリアに閉じたトラフィック量の割合
- ・ 各エリアの割合

##### ○2次元で表現

時間の経過に伴って劇的に変化するデータではないことから、アニメーションや3次元の表現ではなく2次元の表現にする。

したがって、「各エリアとも地域性がない」という傾向をひと目で表現するのに最適なのは、【タイプ H'】であるといえる。

#### 4.トラフィックローカリティの可視化 – 完成へ

タイプ H' を整え、完成させる。

東北が上にくるように配置を変更することによって日本地図のような配置になり、感覚的にわかりやすくなった。都道府県の表現についてはこれからの課題でもあるが、要素をできるだけ少なくまとめて8エリアに区分した。凡例は日本地図を使用し、地理情報と配色を一度に見せることができた。タイプ H'では、各エリアに閉じているトラフィックを各エリアの数値を取って矩形で表していたが、重なってわかりづらいため、平均（黒の太線）の数値を取った。また、図がどのように構成されているのかをマトリクスを添えて説明している。図大きくみせることを最優先し、説明文を加え、配色やレイアウト、フォントやサイズなど見やすく仕上げで完成とした。（最終頁）

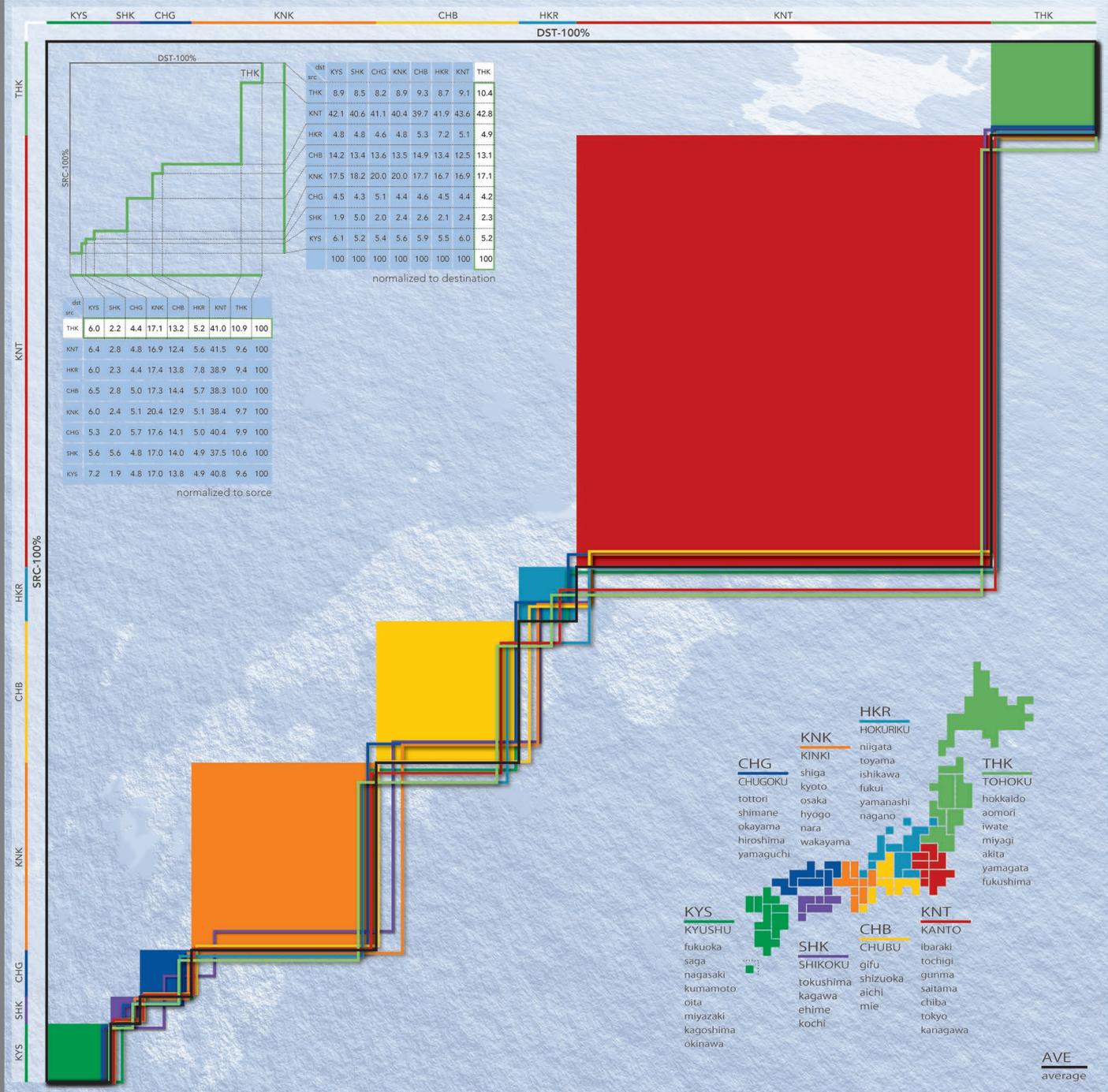
#### 5.まとめ

「何を言いたいか」は、可視化するために最も重要なポイントである。今回は、これを見つけるために試行錯誤を重ねて多くの時間を費やした。「何を言いたいか」が明確であれば、「phase3 どのような手法にするか」、「phase4 どのように見せるか」に早い段階でたどり着くことができ、より質の高い効果的な可視化、さらにはプレゼンテーションが可能になる。そのためには「何が見えるか」に留まらず、そこから「何を言いたいか」「どう見せたいか」を積極的に考えていくべきだろう。日頃から何を言おうとしているのか意識して物事を見ることが、可視化やプレゼンテーションをする際に大きな助けとなってくれるはずである。

今後も継続して取り組み、「可視化マニュアル」を作れるくらいに事例を積み重ねて、プレゼンテーション能力の向上につなげていきたい。

Copyright Notice

Copyright (C) WIDE Project (2006).All Rights Reserved.



# Geographic Traffic Matrix of Japanese User-to-User Traffic

The graph visualizes the regional traffic matrix of Japanese domestic user-to-user traffic. In this graph, step functions are used instead of a standard matrix in order to emphasize the similarity among traffic distributions of different regions.

The measured traffic is in volume (byte count) and covers domestic residential broadband users.

Each step function shows the percentages of source and destination regions of the traffic destined to and originated from the corresponding region.

The step functions look similar for all the regions, and the distribution is roughly proportional to the populations of the regions. The results show that residential users communicate with users in other regions regardless of the region he or she lives in. That is, Internet communication has poor locality, and differs from telephone communication where users tend to talk to nearby neighbors.

The colored rectangular regions represent the regional traffic volumes where both sources and destinations belong to the same region. The size of each rectangle is roughly proportional to the total volume of the regional traffic.

## Methodology

The data was collected at edge routers accommodating residential broadband customers from one domestic ISP using Sampled Netflow.

The regional traffic volumes are obtained by mapping IP addresses to prefectures using Cyber Area Research Inc.'s SURFPOINT Geo-IP database.

The database covers residential customers of domestic ISPs but does not cover domestic leased-lines, data-centers, and international address blocks. Thus, the data used in the graph covers only domestic user-to-user traffic.

The step functions are derived from a standard traffic matrix, a matrix from ingress regions to egress regions.

The distribution in the matrix is normalized to the sources (row) and the destinations (column) independently, and then, a step function is created to show the regional percentages destined to and originated from the region. The top-left graph shows an example: how a step function for Tohoku region is derived from normalized traffic matrices.

