



HAL
open science

L'impact du voisinage géographique des pays dans l'attribution des votes au Concours Eurovision de la Chanson

Jean-Francois Gleyze

► **To cite this version:**

Jean-Francois Gleyze. L'impact du voisinage géographique des pays dans l'attribution des votes au Concours Eurovision de la Chanson. 2009. halshs-00420416

HAL Id: halshs-00420416

<https://shs.hal.science/halshs-00420416>

Preprint submitted on 29 Sep 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

L'impact du voisinage géographique des pays dans l'attribution des votes au Concours Eurovision de la Chanson

Jean-François Gleyze (Dr)

Institut Géographique National – Service de la Recherche

73 avenue de Paris – 94165 Saint-Mandé cedex

e-mail : jf.gleyze@gmail.com

Résumé

Lors de chaque édition annuelle du Concours Eurovision de la Chanson, les pays compétiteurs présentent tour à tour une chanson originale, puis votent en attribuant des points aux pays dont ils ont aimé la chanson. Le pays recueillant le plus de points est alors déclaré vainqueur.

Bien que la qualité des chansons soit théoriquement le seul critère pris en compte par les jurys dans l'attribution des points, l'observation des votes laisse supposer que d'autres critères interviennent – dont notamment la proximité géographique des pays.

Afin de confirmer ou d'infirmer cette assertion, nous élaborons dans cet article une méthode permettant d'analyser les votes échangés entre les pays en regard de leur organisation spatiale.

Pour cela, nous identifions dans un premier temps les couples de pays ayant une tendance significative à attribuer des votes anormalement élevés au cours des éditions de la période 1993 – 2008. Grâce à un test statistique, nous montrons alors que les « sur-votes » ainsi identifiés entre les pays peuvent significativement être expliqués par un critère de proximité géographique.

L'analyse du graphe des sur-votes permet par ailleurs de mettre en évidence l'existence de trois blocs de pays sur-votant massivement les uns pour les autres. Nous montrons enfin que ces blocs structurent fortement les votes anormalement élevés attribués au cours de l'édition 2009.

Cette méthode d'analyse conjointe d'un graphe social et d'un graphe géographique peut être adaptée à l'étude de phénomènes mettant en scène des relations entre des individus spatialisés.

Mots-clés

réseau social, graphe de voisinage spatial, clustering, Concours Eurovision de la Chanson

Introduction

Le Concours Eurovision de la Chanson (Eurovision Song Contest, ou « ESC ») se tient chaque année depuis 1956. Ce concours se déroule en deux temps :

- dans un premier temps, les pays compétiteurs présentent tour à tour une chanson originale,
- puis dans un second temps, les pays votent en distribuant des points pour les pays dont ils ont aimé la chanson.

L'attribution des points telle qu'on la connaît aujourd'hui date de 1975 : chaque pays exprime une dizaine de votes valant, dans l'ordre décroissant, 12, 10, 8, 7, 6, 5, 4, 3, 2 et 1 point(s). À l'issue des votes, le pays recueillant le plus de points est déclaré vainqueur de l'édition.

Le nombre de pays participants est en constante progression depuis la création du concours. Les pays fondateurs, au nombre de sept (Bénélux, France, Suisse, Italie, Allemagne de l'Ouest) ont été rapidement rejoints par de nombreux pays membres de l'Union Européenne de Radio-télévision. Dans les années 70 et 80, une vingtaine de pays concourent ainsi annuellement. L'éclatement du bloc soviétique et de la Yougoslavie au début des années 90 porte ce nombre à 25 en 1993. Aujourd'hui, une quarantaine de pays postulent chaque année pour concourir lors de la phase finale du jeu.

Bien que les pays soient invités à voter pour leurs chansons préférées, les media signalent depuis plusieurs années que l'attribution des points entre les pays pourrait être orientée par la proximité géographique des pays (Wikipedia, 2009a & 2009b). Cette tendance est d'autant plus décriée que le public intervient par télévote dans l'attribution des points depuis 1997.

Dans ce contexte, nous proposons de mettre au point une méthodologie permettant, à partir des votes, d'identifier les liens privilégiés entre les pays (les « sur-votes »), puis de confronter ces liens avec la carte d'Europe afin de démontrer l'existence d'un effet de voisinage dans l'attribution des points. Le cadre de cet étude est exclusivement quantitatif et, à défaut d'apporter des réponses sur les facteurs musicaux, culturels ou politiques intervenant dans l'attribution des votes à l'ESC, met en évidence les structures émergentes du réseau social des votes et du réseau géographique des pays qui peuvent faciliter l'étude de ces facteurs.

Dans une première partie, nous présentons le cadre de notre étude : nous précisons les modalités de l'ESC, nous situons la problématique de notre travail par rapport aux études antérieures et nous décrivons les données de notre étude, tant d'un point de vue contextuel (le réseau social des votes) que d'un point de vue géographique (le réseau de voisinage des pays).

Dans une deuxième partie, nous établissons une méthode statistique d'analyse des votes. Pour chaque couple de pays (i, j) et pour chaque édition, nous considérons le nombre de points que i attribue à j . Compte-tenu du score final obtenu par j lors d'une édition donnée, nous évaluons par un test statistique si ce vote s'écarte significativement du vote médian ou non. Par agrégation statistique sur toutes les éditions considérées, la série des votes de i à j permet de déterminer si le pays i révèle une tendance significative à exprimer des votes anormalement élevés (« sur-votes ») ou des votes anormalement faibles

(« sous-votes ») à l'égard du pays j , indépendamment de la qualité de la prestation du pays j .

Dans une troisième partie, nous analysons le graphe des « sur-votes » et le graphe des « sous-votes » ainsi obtenus sur la période 1993-2008. Compte-tenu de la complexité de ces graphes, nous procédons au préalable à une analyse cartographique afin de comprendre l'organisation de ces votes singuliers. Les cartes obtenues font état de nombreux sur-votes entre pays voisins ; *a contrario*, les sous-votes identifiés concernent majoritairement des pays géographiquement éloignés les uns des autres. En confrontant le graphe des sur-votes et le graphe de voisinage des pays, nous démontrons que les sur-votes concernent significativement des pays géographiquement proches et validons ainsi l'hypothèse du lien présumé entre sur-vote et voisinage géographique. L'analyse du graphe social construit sur les sur-votes permet par ailleurs d'identifier trois blocs de pays ayant une forte tendance à sur-voter les uns pour les autres.

Dans une quatrième et dernière partie, nous analysons l'édition 2009 de l'ESC à la lumière des résultats obtenus sur l'ensemble des 16 éditions précédentes. La méthode statistique que nous avons élaborée dans la deuxième partie permet d'identifier les votes anormalement élevés lors de cette édition : ceux-ci confirment les tendances dégagées précédemment (sur-votes entre pays voisins, sur-votes intra-blocs) et suggèrent de possibles extensions pour les trois blocs identifiés.

L'ensemble des techniques développées dans ce travail peuvent être généralisées à l'analyse comparative de réseaux de votes et de réseaux de voisinage en géographie.

Cadre et problématique de l'étude

Le Concours Eurovision de la Chanson – Pays participants et système de votes

L'analyse statistique des votes échangés lors des nombreuses éditions de l'ESC est une entreprise complexe, car le contingent des pays admis à concourir change significativement d'une année sur l'autre¹. Cette fluctuation est accentuée par l'extension continue du contingent des pays candidats (et regroupés au sein l'Union Européenne de Radio-télévision).

En pratique, le nombre de participants a beaucoup augmenté depuis la création de l'ESC, en même temps que de nouveaux pays sont apparus sur la carte d'Europe avec la chute du bloc soviétique. Les années 1993-1994 constituent à ce titre des dates-clés, car elles voient l'arrivée de dix nouveaux pays candidats (l'Estonie, la Lituanie, la Pologne, la Russie, la Slovaquie, la Roumanie, la Hongrie, la Bosnie-Herzégovine, la Slovénie et la Croatie), suivis peu de temps après par la Macédoine (1998) et la Lettonie (2000). Depuis 2003, plus d'une dizaine de pays ont rejoint l'ESC – pour la plupart situés dans l'ex-Yougoslavie, l'Europe de l'Est et le Caucase (cf Figure 1).

¹ Les votes des éditions étudiées (1993-2009) sont extraits du site <<http://www.kolumbus.fi/jarpen/>>. Les informations générales relatives au concours sont issues de (Wikipedia, 2009a & 2009b).

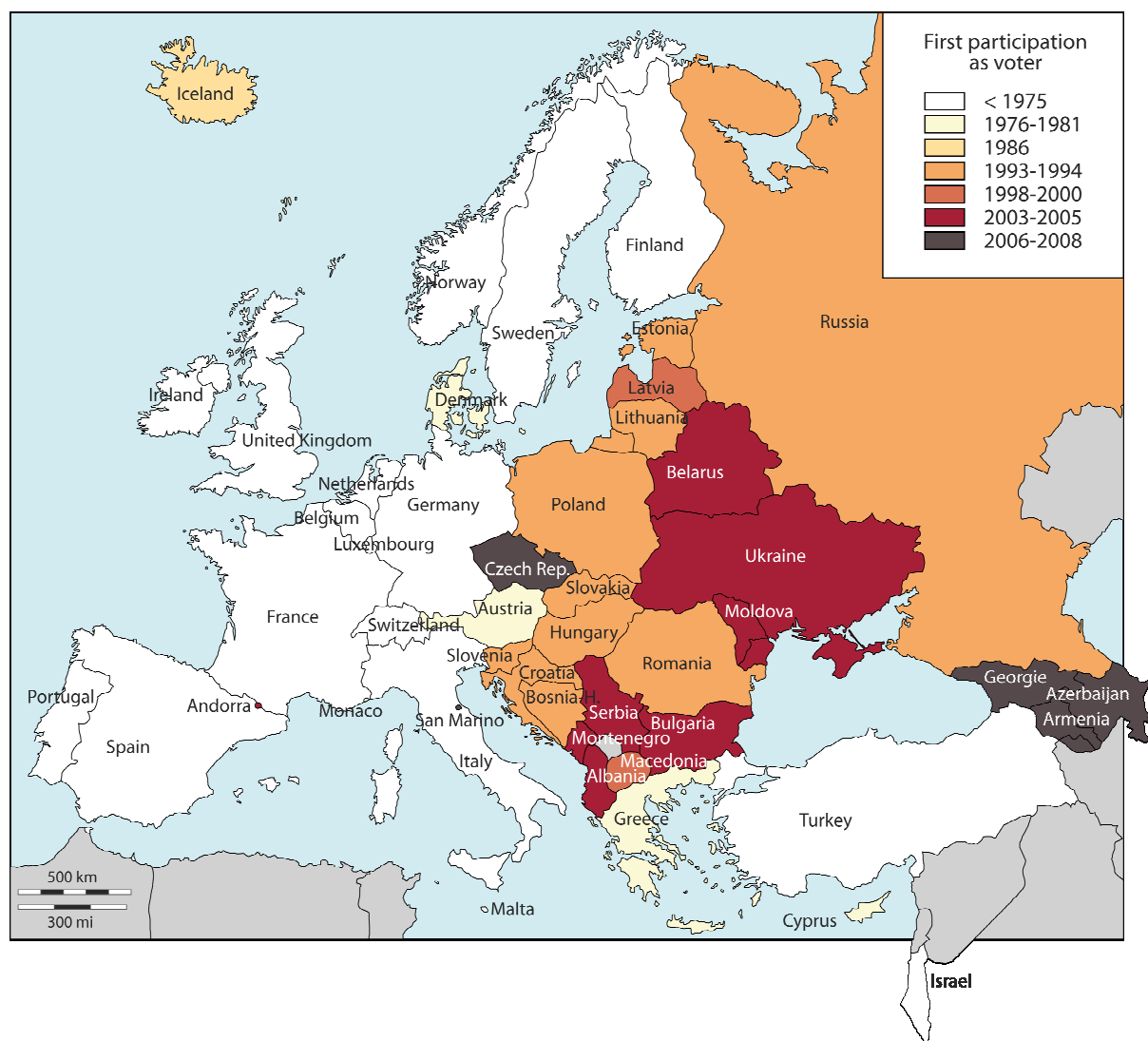


Figure 1 : date d'entrée des pays dans l'ESC²

L'attribution des points par les pays est une procédure très longue, car les jurys de chaque pays interviennent chacun leur tour à l'antenne pour proclamer leurs votes. Au fil des années, les règles ont été adaptées afin de limiter le nombre de pays en compétition.

Ainsi, de 1975 à 2003, le nombre de pays admis en finale a varié de 18 à 26 : lors de ces éditions, chaque pays présentait une chanson et distribuait dix votes de 1 à 12 comme décrit plus haut.

Depuis 2004, des demi-finales ont été mises en place afin de sélectionner les pays finalistes. Entre 2004 et 2009, 36 à 42 pays postulent annuellement pour accéder à la finale, et parmi eux, seuls 24 ou 25 pays sont admis en finale. Lors des finales de ces éditions, tous les pays (demi-finalistes et finalistes) sont invités à distribuer leurs votes aux pays finalistes.

Malgré la grande fluctuation des pays compétiteurs et votants depuis les débuts de l'ESC, les règles de votes n'ont pas changé depuis 1975. À la fin des

² Le fond de carte de l'Europe est issu de la cartothèque du site Histoire-Géographie de l'Académie Aix-Marseille, © Daniel Dalet
<<http://histgeo.ac-aix-marseille.fr/carto/index.htm>>

prestations des pays compétiteurs, le jury de chaque pays votant³ attribue un contingent de points aux dix pays dont il a préféré les chansons (12 points pour le premier, 10 points pour le deuxième, puis de 8 à 1 point(s) pour les pays qu'il classe de la troisième à la dixième place) – étant entendu qu'un pays votant ne peut pas s'attribuer de points.

Problématique – La mise en évidence d'un effet géographique dans l'expression des votes

Fondamentalement, l'attribution des points est dictée par le jugement que portent les différents jurys sur la qualité des prestations des pays compétiteurs. La masse des votes⁴ observés depuis 1975 permet d'envisager une analyse statistique de ces votes et de vérifier si d'autres critères que la qualité interviennent dans l'attribution des points.

Sur la base de ce constat, plusieurs études récentes ont démontré sous certaines hypothèses :

- l'impact de proximités culturelles et linguistiques dans l'attribution des points (Ginsburgh & Noury, 2004) ;
- la similarité des comportements de votes de certains pays (Fenn *et al.*, 2005) ;
- et enfin l'existence de votes collusoires entre certains couples de pays (Fenn *et al.*, 2005, Gatherer, 2006) conduisant à l'émergence de blocs (*op. cit.*, Yair, 1995, Dekker, 2008).

Dans ces études, l'effet du voisinage géographique sur la propension à attribuer des points est constatée pour certains couples de pays (Fenn *et al.*, 2005, Gatherer, 2006) ou mise en évidence ponctuellement pour une édition donnée (Dekker, 2008). Pour autant, aucune des études citées ne démontre l'existence d'un tel effet, aussi bien à des échelles (spatiales et temporelles) locales que globales.

En analysant les points échangés entre les pays au cours des éditions de l'ESC, notre objectif dans cet article est de démontrer que la proximité géographique est un critère favorisant l'attribution de points d'un pays à un autre. Pour ce faire, nous proposons d'identifier statistiquement les votes anormalement élevés entre les pays sur une large période et de les confronter au graphe de voisinage des pays d'Europe. Dans la lignée des travaux cités ci-dessus, nous proposons par ailleurs d'exploiter ces résultats pour mettre en évidence l'émergence de blocs de pays ayant une forte propension à voter les uns pour les autres. Nous proposons enfin de valider les blocs ainsi identifiés en les confrontant aux votes anormalement élevés de la dernière édition en date (2009).

³ La constitution des jurys a néanmoins changé au fil des années. En particulier, depuis 1997, les téléspectateurs peuvent télévoter et influencer le vote émis par le jury de leur propre pays. Pour autant, les règles régissant l'attribution des points sont inchangées depuis 1975.

⁴ soit : 19572 votes échangés entre 51 pays pour les 35 éditions de 1975 à 2009

Description des données d'étude – Votes et voisinages géographiques

Afin de travailler avec un ensemble d'individus statistiques (*i.e.* les pays) globalement homogène, nous proposons de nous restreindre aux 16 éditions de la période 1993-2008. Sur cette période, on dénombre 48 pays potentiellement participants. Parmi eux, seuls 23 à 26 pays sont appelés à concourir annuellement, auxquels s'ajoutent 12 à 18 votants (mais non compétiteurs) pour les éditions incluant une présélection (*i.e.* de 2004 à 2008) (cf Tableau 1).

Tableau 1 : nombre de pays compétiteurs (c) et nombre de pays votants (v) lors de chacune des éditions de l'ESC entre 1993 et 2009

	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09
C	25	25	23	23	25	25	23	24	23	24	26	24	24	24	24	25	25
V												36	39	38	42	43	42

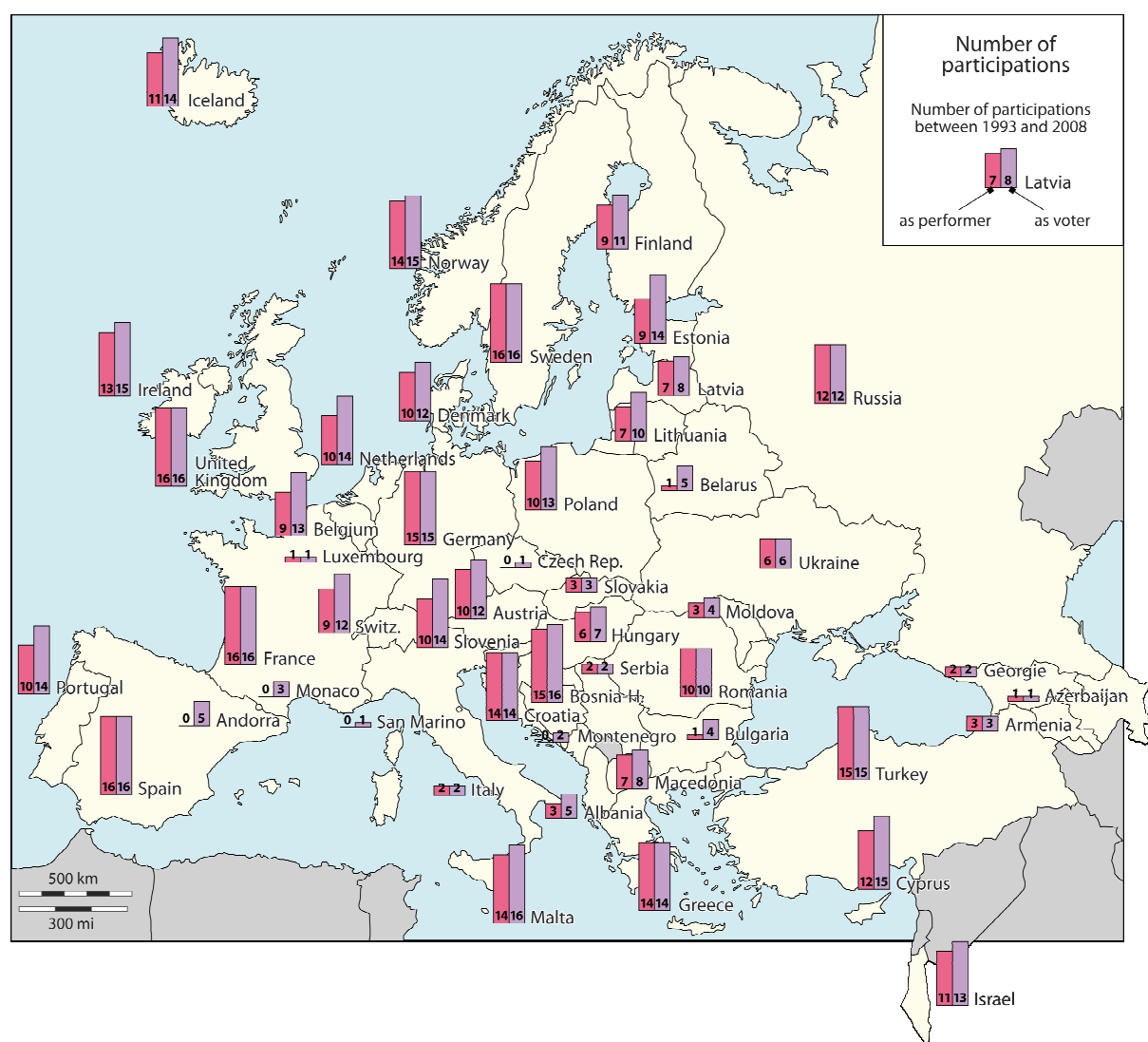


Figure 2 : nombres de fois où les pays ont participé en tant que pays votant (bâtons violets) et en tant que pays compétiteur (bâtons rouges) entre 1993 et 2008

La Figure 2 détaille pour cette même période les nombres de fois où chacun des pays a concouru en tant que pays votant et en tant que pays compétiteur.

Dans la suite de notre travail, nous proposons d'analyser les votes par couples de pays (i, j) (i étant l'émetteur du vote, et j le récepteur de ce vote). Cependant, le nombre restreint de pays compétiteurs à chaque édition (seulement la moitié environ des 48 pays concourent en finale) réduit le nombre total d'éditions où le pays i est amené à exprimer un vote (même nul) pour j . De fait, plus de la moitié des 2256 couples de pays possibles ne se sont trouvés réunis que 0, 1, 2 ou 3 fois sur l'ensemble des 16 éditions (cf Tableau 2). Comme nous le verrons en deuxième partie, cette disparité dans les occurrences des votes d'un pays i vers un pays j ne compromet pas l'analyse statistique du comportement de i à l'égard de j , si ce n'est que les comportements anormaux (« sur-votes » ou « sous-votes ») sont plus difficilement détectables pour les couples de pays dont on dispose de peu d'observations.

Tableau 2 : distribution des 2256 couples possibles de pays (i, j) (i votant, j compétiteur) en fonction du nombre d'éditions où ceux-ci apparaissent (parmi les 16 éditions de 1993 à 2008)

N	0	1	2	3	4	5	6	7	8
n	349	356	300	230	65	92	129	104	89
N	9	10	11	12	13	14	15	16	
n	115	109	62	63	70	66	37	20	

En pratique, nous travaillons donc sur un tableau de données de dimensions $48 \times 48 \times 16$:

- les deux premières dimensions se rapportent aux pays émetteurs et récepteurs des votes (parmi les 48 pays participants indexés pour notre période d'étude) ;
- la troisième dimension se rapporte aux éditions (parmi les 16 éditions s'étant tenues entre 1993 et 2008).

La cellule (i, j, k) de ce tableau recense ainsi le nombre de points que le pays i attribue au pays j au cours de l'édition k .

Si, au cours de cette édition, ce vote ne pouvait avoir lieu (soit que le pays i ne votait pas, soit que le pays j ne concourait pas), nous attribuons la valeur -1 à cette cellule : ce point est important, car il nous permet de distinguer *le fait de ne pas donner de points* (valeur 0) et *le fait de ne pas être en mesure de le faire* (valeur -1).

En parallèle des votes, nous établissons le graphe de voisinage des pays participants. Ce graphe correspond au graphe dual de la carte d'Europe (pour plus de détails sur la notion de « graphe dual » en théorie des graphes, voir par exemple Berge, 1973) : les pays sont représentés par des sommets, et les sommets sont reliés entre eux s'ils partagent une frontière.

Cette règle de construction est incomplète pour constituer un graphe connexe, car certains pays sont insulaires (les îles britanniques, Malte, Chypre, l'Islande) tandis qu'Israël est déconnecté des autres pays (car aucun de ses pays voisins ne participe à l'Eurovision).

Nous choisissons donc de faire les choix suivants :

associés (bien que ne construisant pas explicitement de graphe dual, Dekker, 2008, propose une définition équivalente de la proximité géographique entre deux pays X et Y en comptant « le nombre de frontières à traverser pour aller du pays X au pays Y »⁶).

Avec ce graphe, nous pouvons donc établir une matrice 48×48 des proximités géographiques entre les 48 pays participants à l'ESC entre 1993 et 2008.

Avant de procéder à l'étude statistique des votes et à leur confrontation au graphe de voisinage des pays, il est intéressant d'analyser la structure de ce graphe et la position relative des pays en son sein. Traditionnellement, la carte de l'Europe (telle qu'elle se présente en Figure 2) produit une représentation mentale dans laquelle on associe traditionnellement certains pays au « centre de l'Europe » (par exemple l'Autriche ou la Hongrie) et d'autres pays aux extrémités (par exemple la Norvège au nord ou la Russie à l'est). Dans une organisation en réseau, les distances sont déformées par l'agencement des arêtes et les chemins qu'elles créent entre les sommets (cette singularité des organisations spatiales en réseaux est révélée par le principe des anamorphoses). Afin de se faire une idée plus conforme du voisinage relatif des pays participants à l'Eurovision, nous avons cartographié en Figure 4 les éloignements moyens standardisés des pays compétiteurs pour les éditions 1994, 2000, 2005 et 2008 (ces quatre éditions sont associées aux quatre dernières vagues d'entrée de nouveaux pays dans l'ESC – cf Figure 1).

⁶ « [...] the distance D_{XY} between countries [X and Y] [...] measured by the number of borders needing to be crossed in order to travel from country X to country Y » (*op. cit.*)

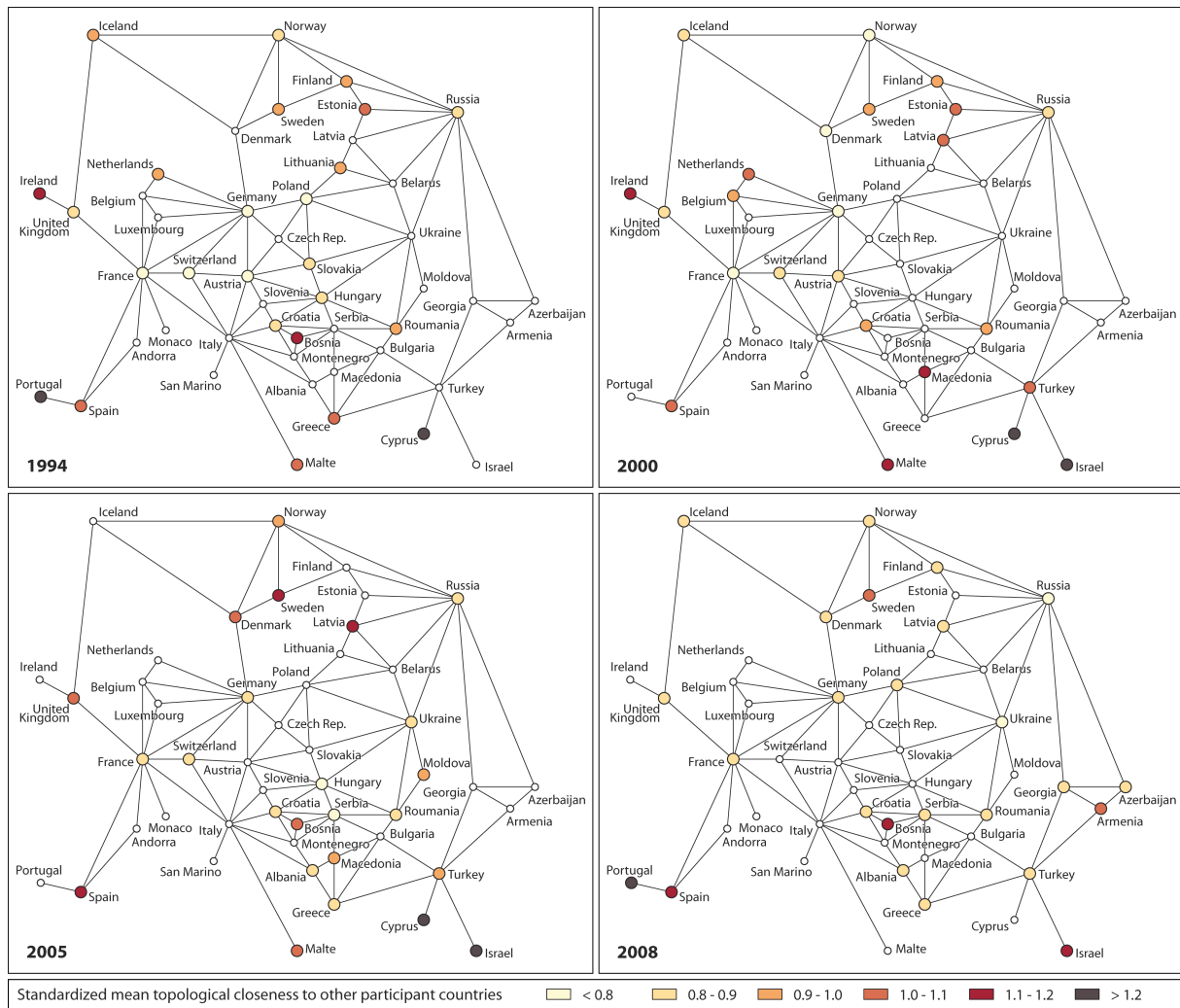


Figure 4 : éloignement moyen standardisé des pays compétiteurs sur le graphe de voisinage (cf Figure 3) pour les éditions 1994, 2000, 2005 et 2008

Pour une édition donnée, l'éloignement moyen d'un pays compétiteur correspond à la distance moyenne (mesuré sur le graphe de la Figure 3) qui le sépare des autres pays compétiteurs. Nous normalisons cette mesure en la divisant par la distance moyenne mesurée sur l'ensemble des paires de pays compétiteurs lors de cette édition – cette standardisation permet de comparer les différentes éditions entre elles (pour plus d'informations sur la notion d'éloignement moyen sur un réseau, voir par exemple Freeman, 1979 ou Wasserman & Faust, 1994). La Figure 4 montre que le « centre » de l'Europe des pays compétiteurs se déplace progressivement d'ouest en est au fil des années et met à mal la représentation traditionnelle centre-périphérie de l'Europe. Nous notons en particulier le positionnement avantageux de la Russie et de l'Ukraine en vertu de leurs nombreuses connexions. Ce positionnement exprime aujourd'hui tout son potentiel alors que la majorité des pays compétiteurs se trouvent dans les Balkans, autour de la mer Baltique et de la mer Noire. Ainsi que nous le verrons dans la suite, il est important de garder en tête ce graphe de voisinage pour analyser la manière dont s'organisent les votes anormalement élevés entre les pays.

Analyse statistique des votes entre 1993 et 2008 – « Sur-votes » et « sous-votes »

Positionnement théorique

Parmi les études antérieures réalisées sur l'ESC, plusieurs d'entre elles se fixent objectif de mettre en évidence des votes anormaux et des ententes entre pays (voir notamment : Fenn *et al.*, 2005, Gatherer, 2006 et Dekker, 2008).

Pour ce faire, ces études doivent s'appuyer sur une description statistique de la situation théorique (ou « situation de référence ») selon laquelle les pays attribuent leurs votes *en fonction uniquement de la qualité des chansons*.

Cette description permet de construire les distributions théoriques des statistiques que l'on souhaite observer sur l'ensemble des votes d'une édition. À un seuil donné, il est alors possible de décider si ces statistiques s'écartent significativement des valeurs théoriques attendues ou non, et donc de déceler des patterns de votes anormaux.

Pour une édition donnée, les distributions théoriques des points attribués d'un pays i à un pays j peuvent être déduites de manière déterministe. Étant donné les règles d'attribution décrites plus haut, le pays i doit répartir dix scores valant respectivement 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 et 12 points parmi m pays compétiteurs. Ainsi, la probabilité qu'un pays compétiteur j reçoive l'un de ces scores en particulier de la part du pays votant i vaut :

- $\frac{1}{m-1}$ si i est également compétiteur (car i ne peut pas voter pour lui-même),
- $\frac{1}{m}$ si i n'est pas compétiteur.

La probabilité qu'un pays j reçoive un score nul vaut :

- $\frac{m-11}{m-1}$ si i est également compétiteur,
- $\frac{m-10}{m}$ si i n'est pas compétiteur.

Telles que nous les avons construites, ces probabilités ne tiennent cependant pas compte de *la qualité des chansons*. Dans ces conditions, la probabilité qu'un pays i attribue 12 points à un pays j est inférieure à $1/m$, quel que soit le couple de pays (i, j) . Ainsi, si l'on souhaite exploiter cette distribution pour décider si un vote de 12 points est anormalement élevé, un tel vote sera effectivement considéré comme tel avec un seuil supérieur à $1/m$, quelle que soit par ailleurs la qualité de la chanson produite par j . Ne disposant d'aucune information complémentaire sur la qualité intrinsèque des chansons, il n'est donc pas possible – en théorie – de rectifier l'appréciation du vote de i à j en tenant compte de la qualité de la chanson de j .

(Fenn *et al.*, 2005) et (Gatherer, 2006) atténuent l'effet de cette lacune en travaillant sur des séries de plusieurs éditions. (Fenn *et al.*, 2005) émettent ainsi l'hypothèse que « tous les pays ont des chances égales de produire des chansons

intrinsèquement 'plaisantes' au cours d'une décennie »⁷. (Gatherer, 2006) étaye ce raisonnement d'un point de vue statistique en remarquant que, « si le vote est réellement suggéré par le mérite artistique de la chanson alors, sur une longue période de temps, les votes de n'importe quel pays doivent approximativement être distribués aléatoirement parmi les autres pays, à compter que les chansons varient en qualité d'année en année et de pays en pays »⁸. Cette hypothèse de travail est acceptable « faute de mieux », mais elle ne tient pas compte explicitement de la qualité des chansons et ne permet pas une analyse des votes à l'échelle d'une seule édition.

A contrario, (Dekker, 2008) analyse exclusivement les votes sur une seule édition et intègre à son analyse la qualité *perçue* des chansons : il propose de définir la qualité perçue d'une chanson par le score final qu'elle obtient à l'issue de l'édition. À défaut de mesurer précisément la qualité intrinsèque de la chanson, cet indicateur mesure pour le moins sa popularité⁹.

Dans le même esprit, nous proposons de conditionner les probabilités d'attribution des points de i à j par le score final obtenu par j pour décrire la situation de référence. De cette manière, un vote de 12 points sera apprécié différemment, selon que le pays destinataire obtienne un score final faible ou élevé.

Notre démarche statistique est originale, car d'une part elle se fonde sur l'analyse individuelle des votes compte-tenu de la qualité perçue des chansons à l'échelle d'une édition donnée, et d'autre part elle permet – par agrégation des probabilités sur une série d'éditions – de mettre en évidence des patterns de votes anormaux entre les pays.

Détermination de la distribution théorique d'un vote et d'un score final

Dans la situation de référence d'attribution des votes, seule la qualité des chansons oriente le choix des jurys. À défaut de disposer d'informations sur la qualité intrinsèque des chansons, nous avons décidé de mesurer celle-ci par le score final qu'elles obtiennent à l'issue de l'édition.

La distribution théorique d'un vote d'un pays i vers un pays j est donc décrite par les probabilités conditionnelles $P(V_{ij}=v \mid S_j=s)$ que le pays j obtienne un score V_{ij} égal à v ($\forall v \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 10, 12\}$) de la part de i , sachant que le score final S_j de j vaut s .

Pour calculer ces probabilités, nous avons besoin de faire l'hypothèse que les votes des jurys sont indépendants les uns des autres. Cette hypothèse est vraisemblable dans la mesure où les jurys sont isolés les uns des autres lorsqu'ils

⁷ « Assuming that all countries have equal chances of producing intrinsically 'nice' songs over the timescale of a decade [...] » (op. cit.)

⁸ « If voting is really on the artistic merit of the song, then over a long period of time any country's votes should approximate to a random distribution among the other countries, as songs will vary in quality from year to year and country to country. » (op. cit.)

⁹ « Nous pouvons mesurer la qualité (ou tout du moins la popularité) de la performance de la chanson du pays Y par le score total S_Y obtenu par Y . Ce total mesure l'ampleur avec laquelle cette chanson a été plébiscitée par une large audience européenne. » (« We can measure the quality (or at least popularity) of country Y 's song performance by using country Y 's total score S_Y . This total provide a measure of how highly that country's song was rated by an Europe-wide audience. ») (op. cit.)

délibèrent et dans la mesure où ils communiquent leurs votes sans avoir connaissance des votes attribués par les autres jurys.

Le calcul de ces probabilités est long, et nécessite en particulier de distinguer des cas, notamment selon que tous les pays votants sont compétiteurs (éditions antérieures à 2003 inclus) ou que seul un nombre restreint de pays votants sont compétiteurs (éditions postérieures à 2003).

L'ensemble de ces calculs est détaillé et illustré en annexe.

À l'issue de ces calculs, nous sommes en mesure de calculer :

- la distribution théorique du vote V_{ij} dans l'hypothèse d'une attribution aléatoire des points [valeurs de $P(V_{ij}=v)$] ;
- la distribution théorique du score S_j sous la même hypothèse [valeurs de $P(S_j = s)$] ;
- et la distribution théorique du vote V_{ij} conditionnellement au score final S_j de j [valeurs de $P(V_{ij}=v | S_j=s)$]

pour tout couple de pays (i, j) et pour toute édition k .

Analyse d'un vote en fonction de sa distribution théorique

Sous l'hypothèse que le choix des jurys est orienté uniquement par la qualité des chansons, le vote V_{ij} d'un pays i pour un pays j est décrit statistiquement par les valeurs de $P(V_{ij}=v | S_j=s)$.

Nous proposons à présent de confronter la valeur observée v_{ij} d'un vote V_{ij} à sa distribution théorique, afin d'apprécier si elle s'écarte ou non des valeurs attendues sous notre hypothèse de travail.

Pour illustrer notre propos, nous prenons pour exemple l'analyse d'un vote v_{ij} attribué d'un pays i à un pays j obtenant un total de 130 points lors d'une édition à 24 compétiteurs ($n = m = 24$). La Figure 5 présente la distribution théorique de la variable V_{ij} associée sachant que $S_j = 130$.

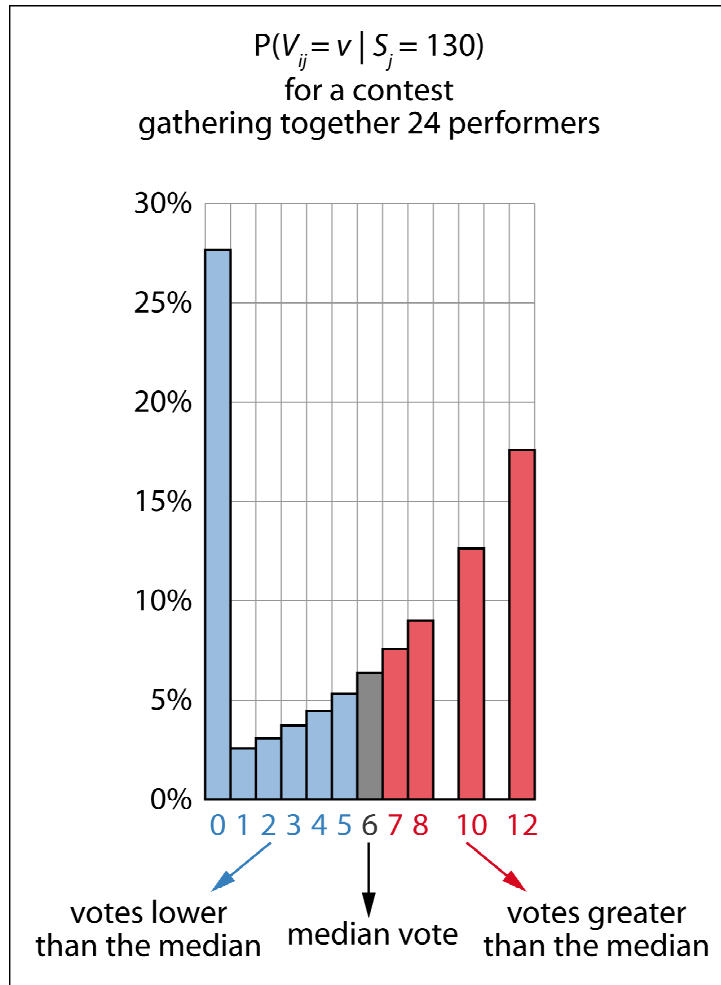


Figure 5 : distribution théorique du vote V_{ij} attribué par un pays i à un pays j sachant que ce dernier obtient un score total S_j de 130, pour une édition rassemblant 24 pays compétiteurs

Sur une telle distribution, on identifie en premier lieu le vote médian m_{ij} . Par définition, il « coupe » la distribution en deux. Il est tel que :

$$\begin{cases} P(V_{ij} \leq m_{ij} \mid S_j = s) \geq \frac{1}{2} \\ \text{et } P(V_{ij} \geq m_{ij} \mid S_j = s) \geq \frac{1}{2} \end{cases}$$

L'identification de ce vote médian permet de repérer les votes strictement supérieurs au vote médian et la probabilité d'obtenir de tels votes :

$$p_{ij}^+ = P(V_{ij} > m_{ij} \mid S_j = s) = \sum_{v > m_{ij}} P(V_{ij} = v \mid S_j = s).$$

Symétriquement, on repère les votes strictement inférieurs au vote médian et la probabilité associée :

$$p_{ij}^- = P(V_{ij} < m_{ij} \mid S_j = s) = \sum_{v < m_{ij}} P(V_{ij} = v \mid S_j = s).$$

Sur l'exemple de la Figure 5, le vote médian vaut 6 points.

La probabilité d'obtenir un vote strictement supérieur à 6 vaut 0,472, et la probabilité d'obtenir un vote strictement inférieur à 6 vaut 0,464.

Dans un deuxième temps, on peut calculer pour chaque vote la p -value qui lui est associée. La p -value est une quantité que l'on utilise pour apprécier la position d'une valeur extrême x au sein de la distribution d'une variable aléatoire X . Ainsi, lorsque la valeur x est élevée, on associe p à la probabilité que la variable aléatoire X soit supérieure à x . Inversement, lorsque la valeur x est faible, on associe p à la probabilité que la variable aléatoire X soit inférieure à x . Dans les deux cas, plus p est faible, plus la valeur x est extrême dans la distribution.

Pour analyser un vote v au sein de sa distribution théorique, on lui associe donc une p -value telle que :

$$\begin{cases} p(v) = P(V_{ij} \leq v \mid S_j = s) \text{ pour } v \text{ inférieur au vote médian} \\ p(v) = P(V_{ij} \geq v \mid S_j = s) \text{ pour } v \text{ supérieur au vote médian} \end{cases}$$

La p -value associée au vote médian vaut 0,5 (la médiane coupant la distribution en deux).

Le Tableau 3 présente les p -values calculées sur la distribution de la Figure 5.

Tableau 3 : p -values associées à la distribution de la variable aléatoire V_{ij} présentée en Figure 5

v	0	1	2	3	4	5	6	7	8	10	12
p	0,277	0,302	0,333	0,370	0,415	0,464	0,5	0,472	0,393	0,302	0,176

Par construction, les p -values servent à identifier les réalisations extrêmes de la variable aléatoire en regard d'un seuil α fixé à l'avance (traditionnellement : $\alpha = 0,1, 0,05$ ou $0,01$). Si la p -value est inférieure à ce seuil, alors on considère que la réalisation est trop excentrée sur la distribution et donc que les hypothèses servant de fondement à cette distribution sont erronées ou incomplètes.

Ainsi, sur notre exemple d'étude, aucun vote ne peut être considéré comme extrême au seuil $\alpha = 10\%$, car aucune des p -values n'est inférieure à ce seuil. Au seuil $\alpha = 20\%$ (seuil trop grossier pour être utilisé en pratique), on considèrerait en revanche que le vote $v = 12$ est anormalement élevé pour la variable aléatoire V_{ij} . Face à un tel vote, on concluerait donc, avec un risque α de se tromper, que la variable aléatoire V_{ij} n'est pas régie par la distribution que l'on a construite, et qu'il existe donc d'autres paramètres que la qualité des chansons intervenant dans l'attribution du vote de i à j .

Appliquée à chacun des votes v_{ij} d'une édition k donnée, cette procédure statistique permet finalement d'identifier au seuil α les votes anormalement élevés et les votes anormalement faibles échangés lors de cette édition.

Analyse d'une série de votes sur plusieurs éditions – Détection des « sur-votes » et des « sous-votes » entre pays

L'identification de votes anormalement élevés ou anormalement faibles est rare à l'échelle d'une seule édition, car les distributions sont telles que les p -values associées aux valeurs extrêmes sont souvent supérieures à 0,1. Ce point est surtout vrai pour le vote nul, car la règle d'attribution des votes est telle que les pays sont amenés à distribuer beaucoup de votes nuls.

Et quand bien même un vote observé v_{ij} au cours d'une édition est identifié comme anormal, il ne peut mettre en évidence à lui seul une anomalie du comportement du pays i à l'égard du pays j . Ainsi que le souligne (Gatherer, 2006), l'ESC est aussi une tribune dans laquelle les pays peuvent exprimer ponctuellement une opinion politique, et ils ne s'en privent pas comme le rappelle l'édition 2003 au cours de laquelle le Royaume-Uni, fraîchement engagé dans la guerre en Irak, n'a récolté aucun point.

Dans ce contexte, il nous apparaît pertinent d'analyser le comportement de vote d'un pays i à l'égard d'un pays j sur une échelle temporelle plus large, en considérant la collection des votes $v_{ij}(k)$ observés sur toutes les éditions k de la période d'étude.

S'agissant de décider si le pays i présente une tendance significative à privilégier – ou au contraire à léser – le pays j , les statistiques que nous choisissons d'observer sur la collection des votes $v_{ij}(k)$ sont :

- le nombre n_{ij}^+ de fois où le vote $v_{ij}(k)$ est strictement supérieur au vote médian $m_{ij}(k)$;
- le nombre n_{ij}^- de fois où le vote $v_{ij}(k)$ est strictement inférieur au vote médian $m_{ij}(k)$.

Afin de décider si ces nombres sont anormalement élevés, on calcule leurs distributions théoriques sous notre hypothèse de travail, et on regarde si leurs p -values sont inférieures à un seuil α fixé à l'avance.

Étant donné que l'on connaît pour chaque édition k la probabilité $p_{ij}^+(k)$ que le vote $V_{ij}(k)$ soit strictement supérieur à la médiane, un simple calcul combinatoire permet de déterminer la distribution théorique du nombre N_{ij}^+ de fois où de tels votes se produisent sur l'ensemble des éditions k (et symétriquement pour le nombre N_{ij}^- de votes strictement inférieurs à la médiane).

À titre d'exemple, le Tableau 4 présente l'historique des votes que la Norvège a attribué à la Grèce sur notre période d'étude. Pour chacun de ces votes, on peut déterminer le vote médian et la probabilité théorique que le vote attribué soit inférieur ou supérieur à ce vote médian, compte-tenu du score final obtenu par la Grèce.

Tableau 4 : historique des votes attribués par la Norvège (i) à la Grèce (j) entre 1993 et 2008 et statistiques associées

édition k	1993	1994	1995	1996	1997	1998	2001	2003	2004	2005	2006	2007	2008
nb votants	25	25	23	23	25	25	23	26	36	39	38	42	43
$S_j(k)$	64	44	68	36	39	12	147	25	252	230	128	139	218
$v_{ij}(k)$	7	4	0	3	5	0	8	0	2	4	0	0	0
$m_{ij}(k)$	0	0	1	0	0	0	7	0	8	7	1	1	5
$p_{ij}^-(k)$	0	0	0,49	0	0	0	0,43	0	0,46	0,49	0,46	0,49	0,46
$p_{ij}^+(k)$	0,45	0,35	0,46	0,34	0,33	0,16	0,49	0,24	0,44	0,43	0,5	0,48	0,49

Ces valeurs nous permettent dans un second temps de construire les distributions théoriques des nombres N_{ij}^- et N_{ij}^+ associés au couple de pays Norvège-Grèce (cf Figure 6).

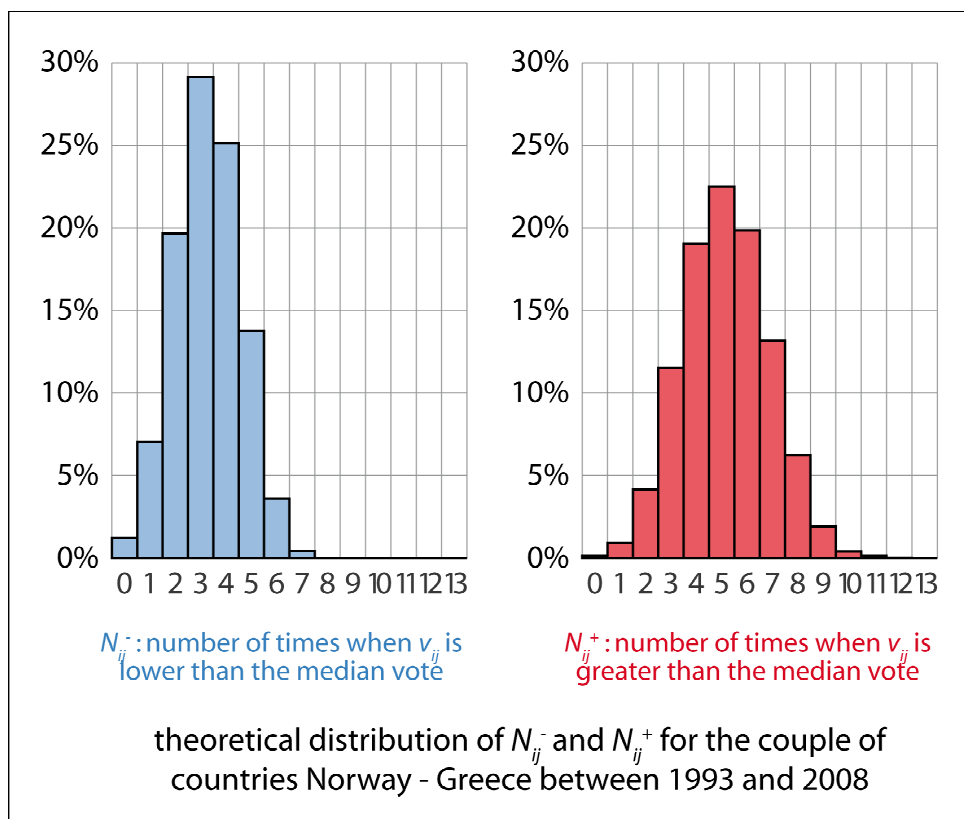


Figure 6 : distributions théoriques des nombres de fois où le vote de la Norvège à la Grèce est inférieur au vote médian (N_{ij}^-) ou supérieur au vote médian (N_{ij}^+) sur la période 1993-2008

Entre 1993 et 2008, nous dénombrons 13 éditions au cours desquelles la Norvège a été amenée à émettre un vote pour la Grèce. Parmi ces 13 éditions, la Norvège a émis 6 fois un vote inférieur à la médiane et 5 fois un vote supérieur à la médiane.

S'agissant de savoir si ces nombres sont significativement élevés, les distributions de la Figure 6 indiquent que les p -values associées valent respectivement 0,0402 et 0,6422. Avec un risque $\alpha = 5\%$ de se tromper, on peut

donc affirmer que la Norvège a une tendance significative à attribuer des votes anormalement faibles à destination de la Grèce.

Appliquée à chacun des couples (i, j) de pays, cette procédure statistique permet d'identifier les tendances qu'ont certains pays à attribuer des votes anormalement élevés (« sur-votes ») et les votes anormalement faibles (« sous-votes ») sur la période 1993-2008.

Nous allons à présent synthétiser l'ensemble des sur-votes et des sous-votes identifiés sur notre période d'étude et procéder à leur analyse.

L'impact de la proximité géographique sur l'attribution des points – Les blocs de pays connivents

Le graphe des sous-votes est de cardinal 25. Cette valeur est faible en regard du nombre total de couples de pays observables sur la période d'étude (1907). Cette faible valeur n'est cependant pas surprenante, car le système d'attribution des points conduit les pays à distribuer un très grand nombre de votes nuls et la détection de votes anormalement faibles en est d'autant plus difficile.

La Figure 7 recense l'ensemble de ces sous-votes.

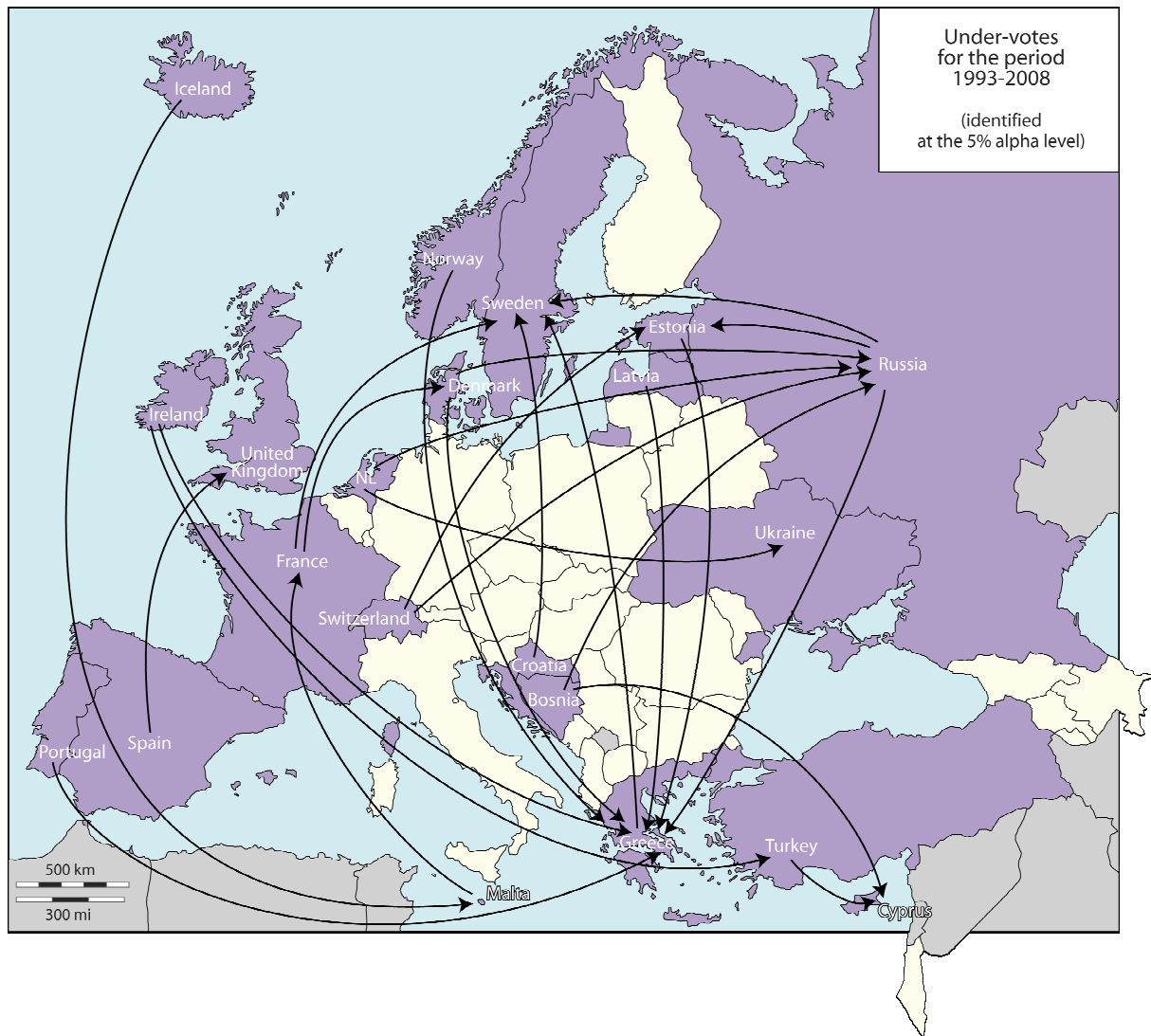


Figure 7 : sous-votes identifiés sur la période 1993-2008 (au seuil $\alpha = 5\%$)

On note sur cette figure plusieurs singularités :

- la majorité des sous-votes sont destinés à la Russie (4), la Grèce (7) et la Suède (4) ;
- il n'existe pas de couples de pays échangeant des sous-votes ;
- enfin, lorsque l'on confronte les sous-votes aux sur-votes, on constate qu'il n'existe qu'un couple de pays associé à un sous-vote dans un sens (Russie → Estonie) et à un sur-vote dans l'autre (Estonie → Russie).

D'un point de vue géographique, il est à noter que seuls deux sur-votes concernent des pays voisins, à savoir Russie → Estonie et Turquie → Chypre – ces liens pouvant être expliqués par les tensions politiques existant entre ces pays.

S'agissant de savoir s'il existe un lien entre sous-vote et voisinage géographique, nous représentons en Figure 8 la distribution des distances entre les 1907 couples de pays d'une part, et la distribution des distances associées aux 25 sous-votes.

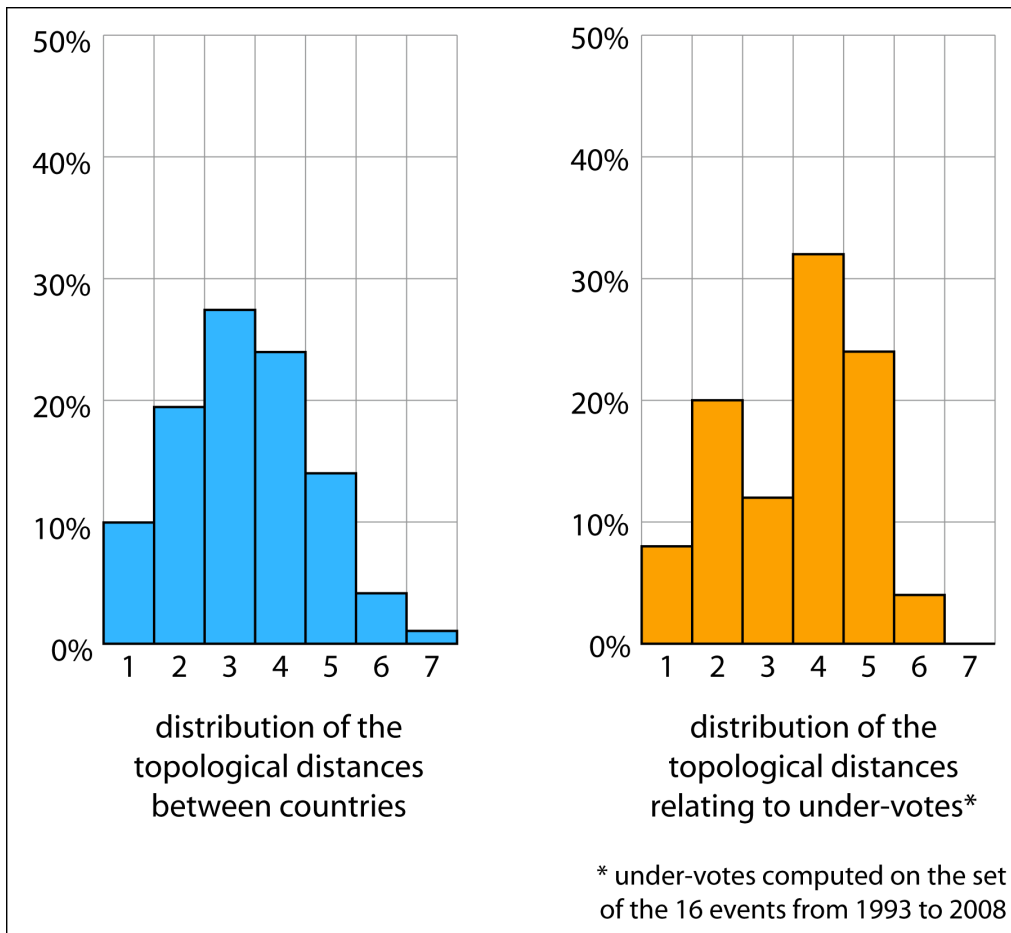


Figure 8 : distribution des distances entre les 1907 couples de pays (à gauche) et distribution des distances associées aux 25 sous-votes (à droite) pour la période 1993-2008

Les 25 sous-votes extraits des 1907 couples de pays présentent une distribution des distances qui semble anormalement excentrée à droite.

Afin de décider si les sous-votes sont significativement associés à des couples de pays éloignés, nous proposons de tester au seuil $\alpha = 5\%$ l'alternative suivante :

- hypothèse H_0 : les sous-votes ne sont pas expliqués par un critère de voisinage ;
- hypothèse H_1 : les sous-votes sont associés à des couples de pays éloignés.

Pour mener ce test, nous prenons comme statistique la moyenne des distances de 25 couples de pays. Pour notre étude, la statistique observée sur les 25 sous-votes vaut 3,560.

S'agissant de déterminer si cette statistique est anormalement élevée (au seuil α), nous mettons en œuvre un « test de permutation » (pour plus de détails sur les tests de permutation, voir par exemple : Good, 1994) : nous extrayons aléatoirement un grand nombre de fois ($n = 10000$) 25 couples de pays parmi les 1907 couples de pays possibles, et pour chacun d'eux, nous mesurons la statistique associée : l'ensemble des n valeurs fournit une distribution approchée de la statistique d'étude sous l'hypothèse H_0 .

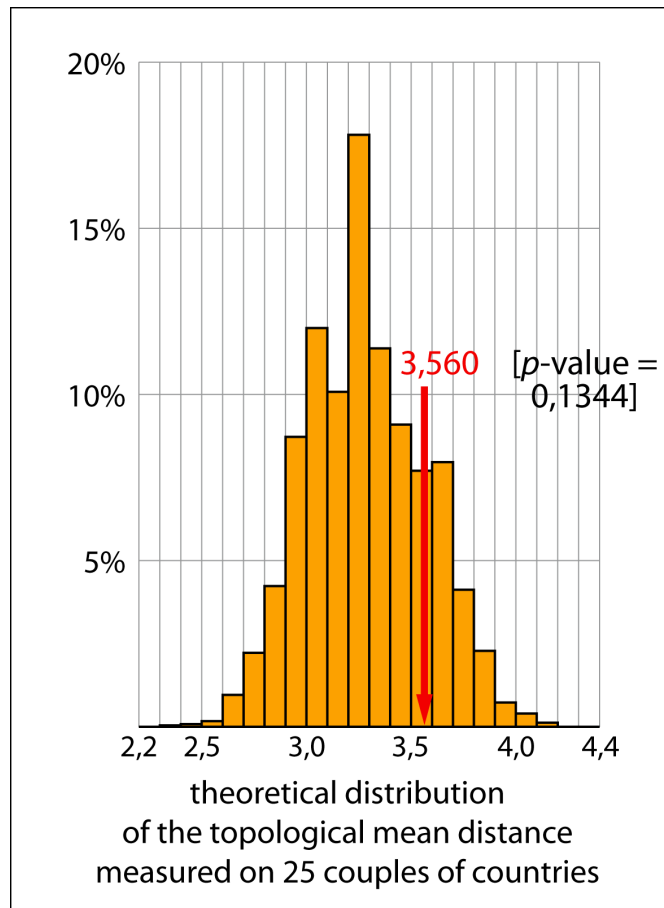


Figure 9 : distribution théorique de la moyenne des distances associées à un ensemble de 25 couples de pays et p -value correspondant à la moyenne des distances associées aux sous-votes

La Figure 9 présente la distribution ainsi obtenue et situe la statistique observée sur les 25 sous-votes. Il apparaît en effet que les sous-votes sont associés à des couples de pays dont la distance moyenne est sensiblement élevée ; néanmoins, la p -value est au-delà du seuil α ($p \approx 13\%$), aussi ne pouvons-nous pas rejeter l'hypothèse H_0 .

Bien que les sous-votes ne peuvent être significativement associés à des couples de pays éloignés, nous allons montrer *a contrario* que les sur-votes concernent significativement des couples de pays proches.

Le graphe des sur-votes est de cardinal 135. Ce nombre est trop élevé pour pouvoir représenter sur une même carte l'ensemble des sur-votes. Néanmoins, afin d'appréhender la morphologie de ceux-ci, nous présentons ci-après les sur-votes émis et reçus de quatre pays, à savoir la Macédoine, la Norvège, la Russie et la Turquie (cf Figure 10).

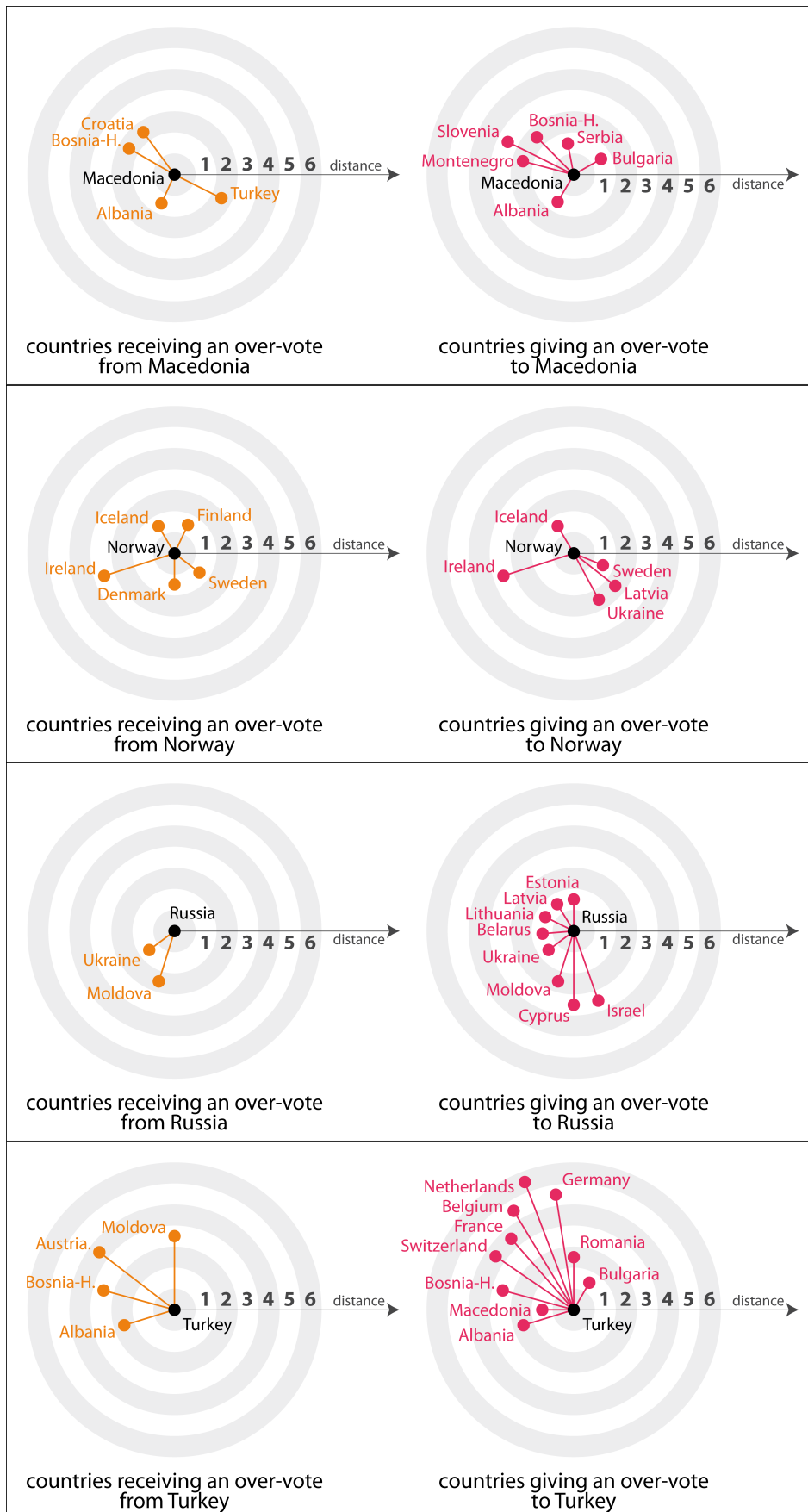


Figure 10 : sur-votes émis et reçus par la Macédoine, la Norvège, la Russie et la Turquie (les cercles concentriques permettent de repérer la distance au pays étudié)

Les exemples que nous avons choisi de présenter en Figure 10 sont représentatifs des tendances que l'on observe lorsque l'on fouille la structure du graphe des sur-votes, à savoir :

- de manière générale, de nombreux sur-votes associant des pays proches,
- et des patterns de sur-votes singuliers, de type politique (par exemple les pays de l'ancienne URSS sur-votent massivement pour la Russie) ou communautaire (les pays accueillant les émigrés turcs votent massivement pour la Turquie).

Par ailleurs, alors que le graphe des sous-votes ne contenait aucun lien réciproque, plus de 40% des sur-votes sont réciproques.

La Figure 11 présente les sur-votes concernant les pays voisins : parmi les 135 sur-votes, 55 sont de ce type (41%), et parmi ceux-ci, 32 sont réciproques (24%).



Figure 11 : sur-votes entre pays voisins, identifiés sur la période 1993-2008 (au seuil $\alpha = 5\%$)

L'importance des sur-votes échangés entre pays voisins et la fréquence de leur réciprocity vont dans le sens d'un lien entre sur-vote et voisinage géographique.

Afin de démontrer formellement ce lien, nous effectuons un test analogue à celui mis en œuvre pour l'étude des sous-votes : nous présentons en Figure 12 la distribution des distances entre les 1907 couples de pays, et nous la comparons à la distribution des distances associées aux 135 sur-votes.

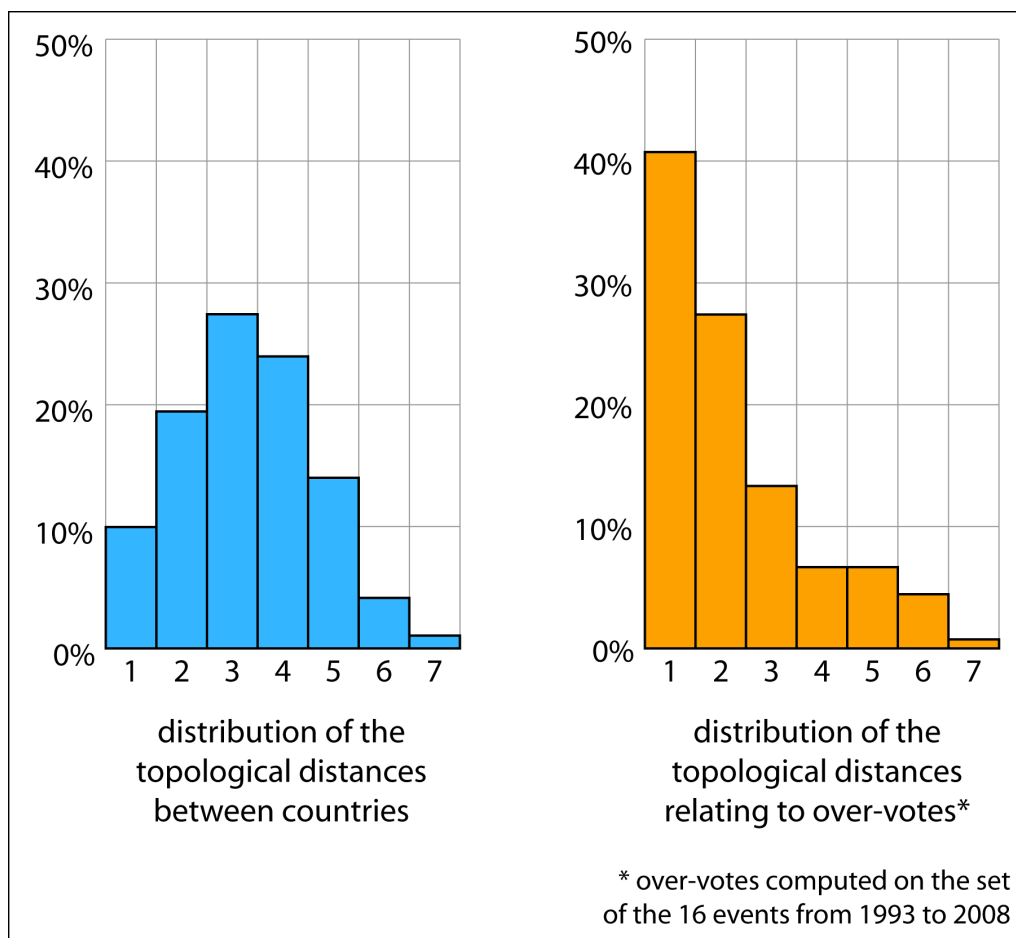


Figure 12 : distribution des distances entre les 1907 couples de pays (à gauche) et distribution des distances associées aux 135 sur-votes (à droite) pour la période 1993-2008

Comme pressenti, les 135 sous-votes extraits des 1907 couples de pays présentent une distribution des distances qui semble anormalement excentrée à gauche.

Nous effectuons donc le test suivant :

- hypothèse H_0 : les sur-votes ne sont pas expliqués par un critère de voisinage ;
- hypothèse H_1 : les sur-votes sont associés à des couples de pays proches.

La statistique associée à ce test est la moyenne des distances de 135 couples de pays. Pour notre étude, la statistique observée sur les 135 sur-votes vaut 2,274. Nous obtenons la distribution théorique de la statistique d'étude par simulation de Monte-Carlo ($n = 10000$).

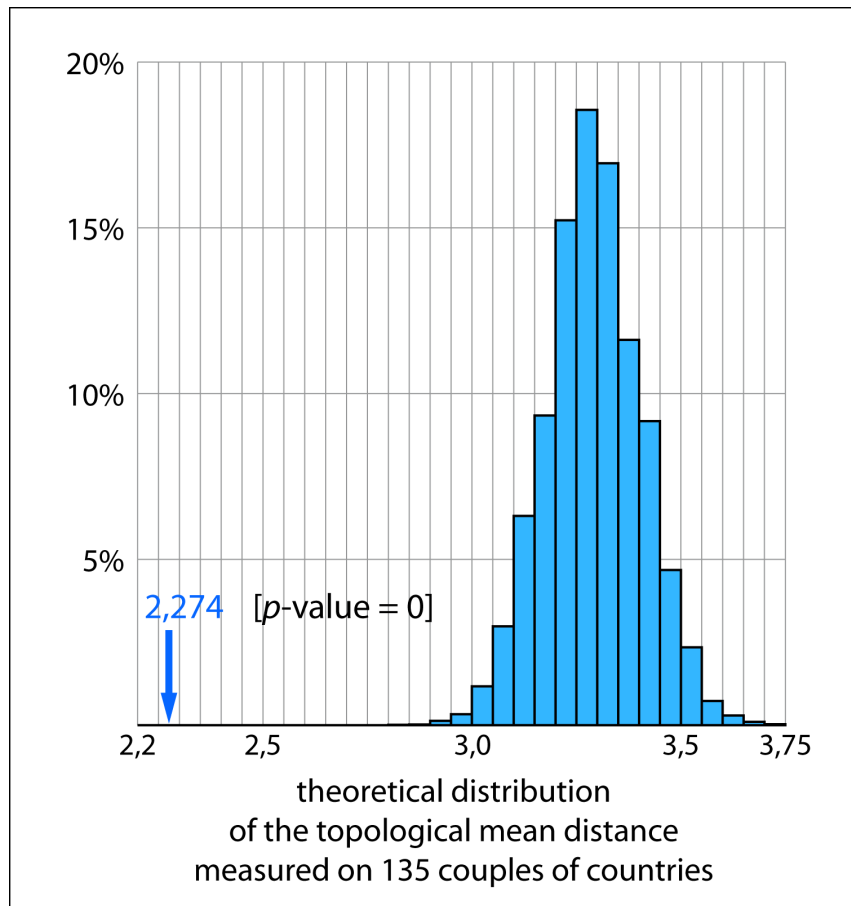


Figure 13 : distribution théorique de la moyenne des distances associées à un ensemble de 135 couples de pays et p -value correspondant à la moyenne des distances associées aux sur-votes

La Figure 13 présente la distribution ainsi obtenue et situe la statistique observée sur les 135 sur-votes : la p -value associée est nulle, aussi rejetons-nous l'hypothèse H_0 .

Ce test nous permet donc de conclure à l'existence d'un lien entre voisinage géographique et attribution de sur-votes entre les pays.

La preuve de ce lien étant établie, nous proposons dans un second temps d'identifier les sous-groupes (ou « clusters ») de pays fortement connectés au sein du graphe des sur-votes et de vérifier si ces sous-groupes correspondent à des secteurs géographiques précis.

Le critère de constitution des sous-groupes le plus strict (et également le plus immédiat à mettre en œuvre) est l'exhaustivité des liens (Wasserman et Faust, 1994) : les sous-groupes ainsi constitués – appelés « cliques » – sont tels que leurs pays membres échangent tous des sur-votes deux à deux, dans les deux directions.

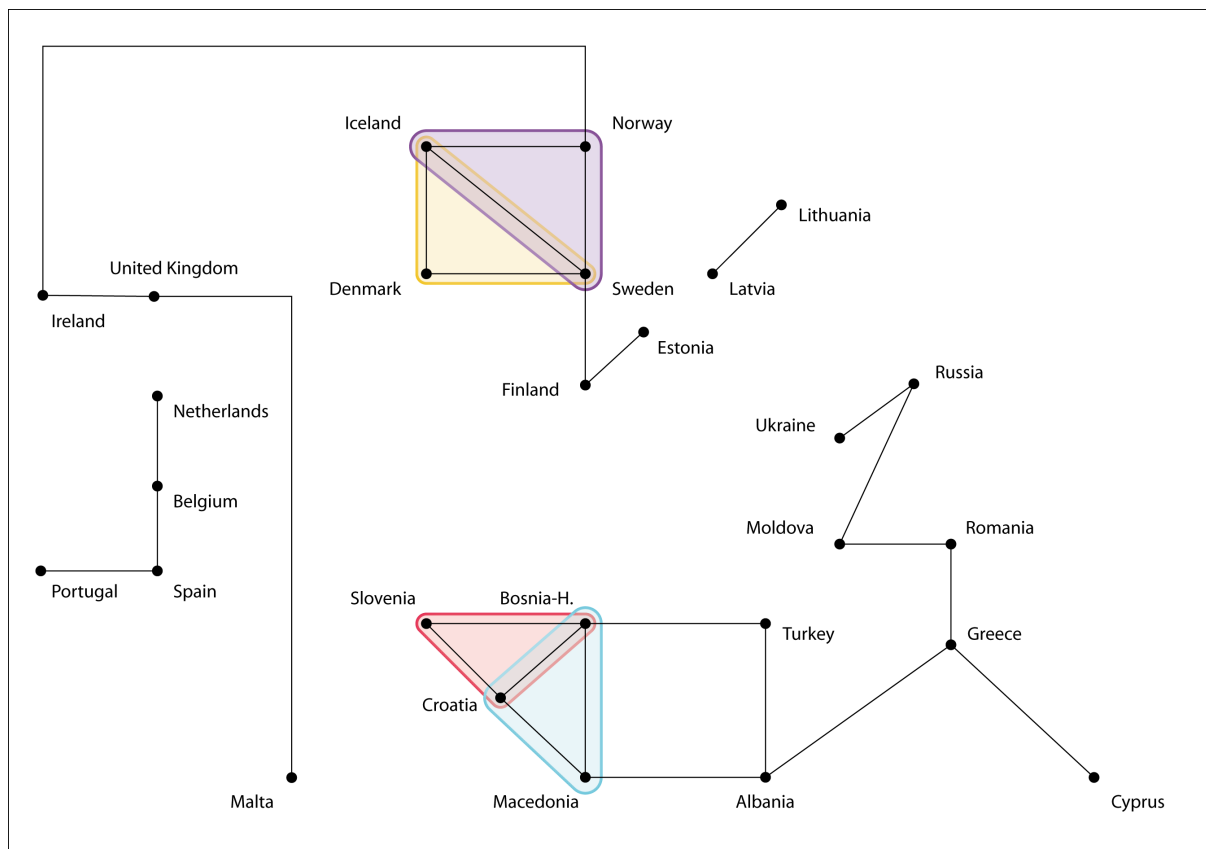


Figure 14 : échanges de sur-votes entre pays (arêtes en noir) et cliques (ensembles de couleur) identifiés sur la période 1993-2008 (au seuil $\alpha = 5\%$)

La Figure 14 met en évidence tous les couples de pays échangeant des sur-votes et les quatre cliques de trois pays construits sur les liens à double-sens. Deux de ces cliques concernent les pays scandinaves et sont articulées sur le tandem Islande – Suède, et les deux autres concernent les pays de l'ex-Yougoslavie et sont articulées sur l'axe Bosnie-Herzégovine – Croatie. Bien que ne pouvant conduire à la formation de cliques, les autres échanges de sur-votes semblent géographiquement très localisés autour de nombreux pôles (pays baltes, pays anglo-saxons, pays d'Europe occidentale, pays méditerranéens, pays d'Europe orientale).

Afin d'identifier des concentrations de sur-votes plus étendues, nous proposons d'atténuer le critère d'exhaustivité dans la recherche des clusters. Au sein d'une clique, la densité des liens doit être égale à 1 (tous les pays sont reliés deux à deux, par un sur-vote réciproque). Nous proposons d'abaisser cette densité à 0,75 et, en contrepartie, de vérifier les deux critères additionnels suivants :

- chaque pays d'un cluster doit être relié à tous les autres pays du cluster par au moins un sur-vote (quel qu'en soit le sens)¹⁰,
- chaque pays d'un cluster émettre au moins un sur-vote et recevoir au moins un sur-vote avec les autres pays du cluster¹¹.

¹⁰ couplé au critère de densité ($\geq 0,75$), ce critère garantit que, au sein d'un cluster, plus de la moitié des couples de pays échangent des sur-votes et qu'il n'existe aucun couple de pays qui ne soit pas relié par un sur-vote

¹¹ ce critère assure que les pays sont effectivement « intégrés » à leur cluster, en émission et en réception

La Figure 15 présente les clusters mis en évidence par ces critères de regroupement.

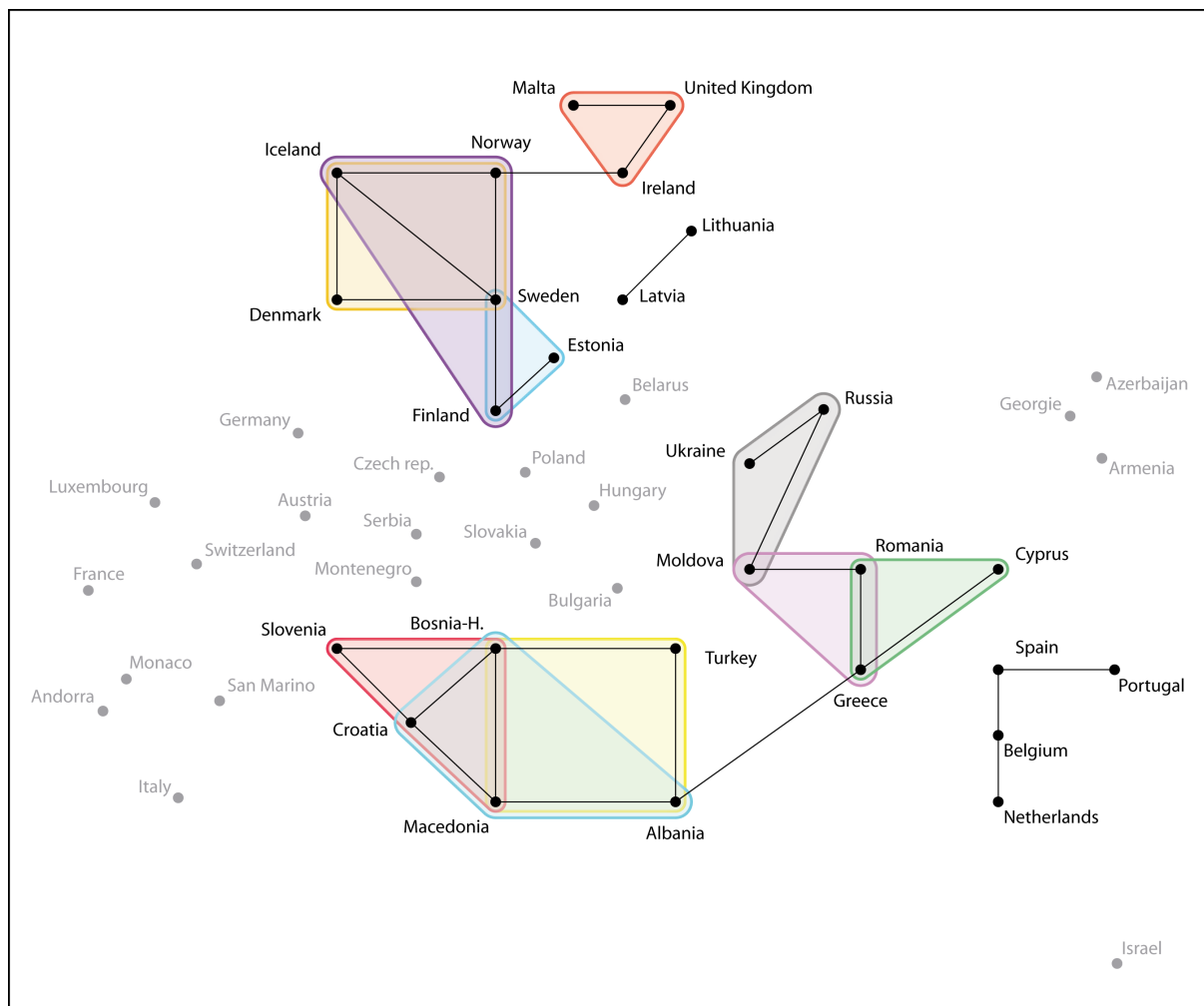


Figure 15 : échanges de sur-votes entre pays (arêtes en noir) et clusters (ensembles de couleur) identifiés sur la période 1993-2008 (au seuil $\alpha = 5\%$) – les clusters sont tels que : 1. la densité des sur-votes est supérieure à 0,75, 2. tout couple de pays est relié par au moins un sur-vote, 3. chaque pays émet et reçoit au moins un sur-vote

Les clusters identifiés affichent une valeur d'éloignement moyen toujours inférieure ou égale à 2 – à l'exception du cluster Chypre – Grèce – Roumanie pour laquelle cette valeur l'excède de peu (7/3). Ces clusters regroupent ainsi des pays géographiquement proches, et leurs recouvrements permettent finalement d'isoler trois grands blocs de pays :

- les pays scandinaves,
- les pays de l'ex-Yougoslavie associés à la Turquie,
- les pays de l'Europe orientale associés à Chypre et la Grèce.

Si l'on étudie dans le détail le graphe des sur-votes, on s'aperçoit que ces blocs sont eux-mêmes renforcés ou étendus par des sur-votes situés en marge :

- ainsi les sur-votes Lettonie → Norvège et Irlande → Lettonie indiquent une possible extension du bloc scandinave aux pays baltes et aux pays anglo-saxons ;
- les sur-votes Chypre → Espagne et Grèce → Espagne d'une part, et les sur-votes Espagne → Moldavie et Espagne → Roumanie d'autre part laissent envisager un rattachement de l'Espagne au bloc d'Europe orientale ;
- enfin les sur-votes Turquie → Moldavie et Roumanie → Turquie établissent des passerelles entre les deux grands blocs emmenés d'une part par la Turquie et d'autre part par la Grèce, pour peut-être aboutir dans les éditions futures à un bloc plus large – mais moins cohésif – autour des Balkans.

Afin de mesurer l'impact que peuvent avoir ces blocs dans les patterns de votes des éditions ultérieures et afin de mettre à l'épreuve les limites et les possibles extensions de ces blocs, nous proposons dans une dernière partie de les confronter aux votes de l'édition 2009.

Comparaison des résultats avec les votes de l'édition 2009 – Perspectives d'étude

Pour caractériser les comportements singuliers des pays dans leur manière de voter en 2009, nous proposons de considérer un à un les votes de cette édition, et de retenir parmi eux les votes se trouvant dans les queues de distribution, c'est-à-dire les votes anormalement élevés et les votes anormalement faibles au seuil $\alpha = 10\%$ (cf deuxième partie).

Les votes anormalement faibles sont au nombre de 3 et sont tous adressés à la Norvège (Bulgarie → Norvège, République Tchèque → Norvège, Turquie → Norvège). Ces informations sont pauvres, car le score obtenu par la Norvège¹² au cours de l'édition 2009 est tel que tout vote pour la Norvège inférieur à 8 points est considéré comme anormalement faible.

En revanche, les votes anormalement élevés sont au nombre de 65 et apportent des informations intéressantes en regard des blocs précédemment identifiés.

Ces votes peuvent être répartis en deux grandes familles :

- d'une part les votes concernant les trois blocs précédemment identifiés,
- d'autre part les votes externes à ces blocs.

Intéressons-nous en premier lieu aux votes anormalement élevés de l'édition 2009 mettant en jeu les trois blocs mis en évidence sur la période 1993-2008 (cf Figure 15). La Figure 16 représentent ces votes en distinguant :

- les votes réciproques,
- les votes non-réciproques internes aux blocs,
- les votes non-réciproques en marge des blocs.

Les votes réciproques et les votes non-réciproques internes aux blocs répondent, sans exception, à des sur-votes de la période 1993-2008. Leur nombre et leur répartition spatiale démontrent la persistance de la structure des blocs et la pérennité de ceux-ci.

¹² 387 points pour un score maximum possible de 492 points

En marge de ces blocs, 8 votes établissent des liens entre pays de blocs différents ou pays marginaux, confirmant les extensions pressenties dans la partie précédente :

- ainsi le cluster Malte – Royaume-Uni – Irlande établit des liens avec les pays nordiques (Malte → Suède et Malte → Finlande) et les pays baltes (Irlande → Lituanie et Royaume-Uni → Lituanie) ;
- la péninsule ibérique émet des votes à destination du bloc est-européen (Espagne → Roumanie et Portugal → Moldavie) ;
- enfin, la perméabilité des clusters des Balkans est confirmée par des votes en faveur de la Roumanie (depuis la Turquie et la Macédoine).

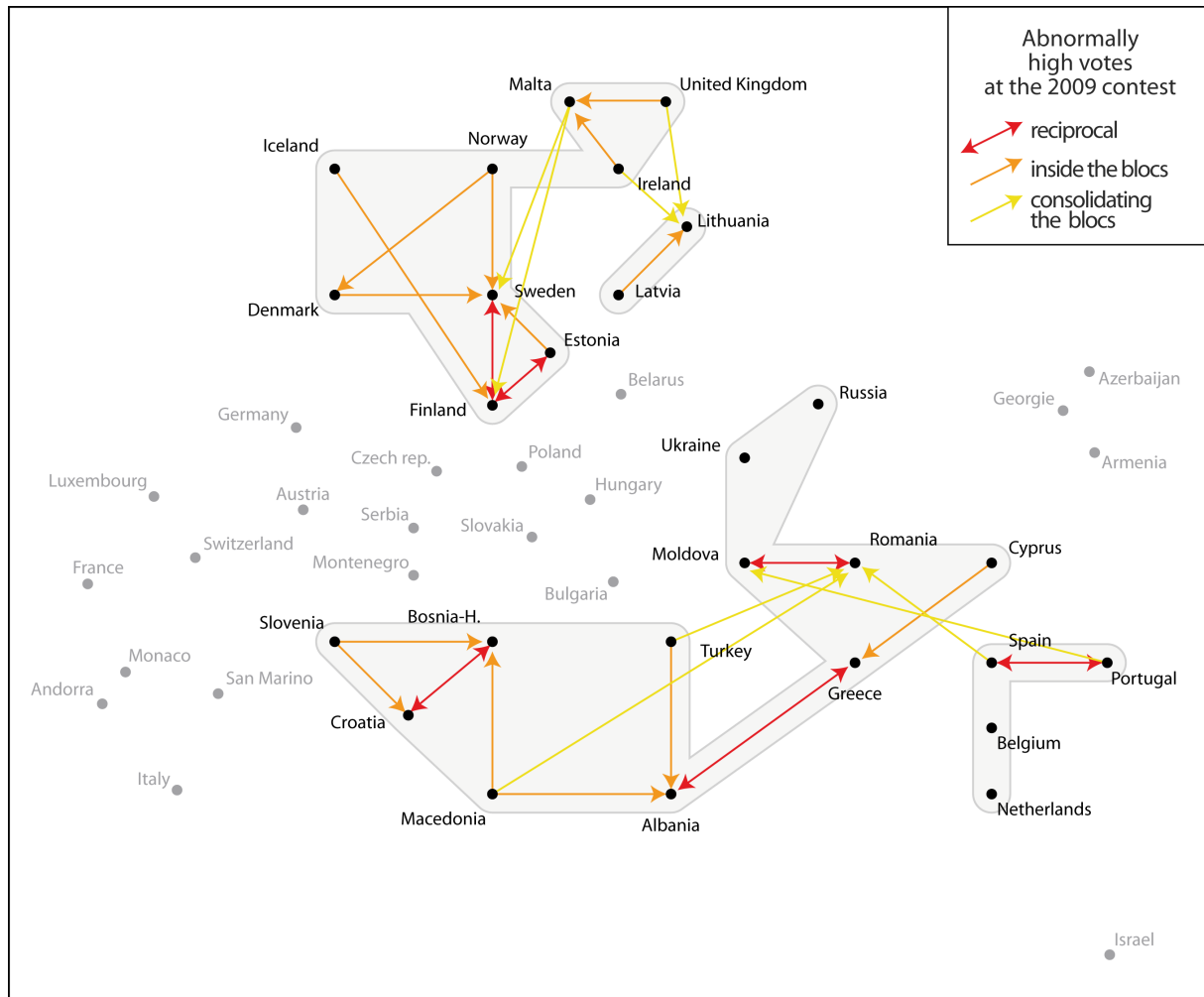


Figure 16 : votes anormalement élevés (au seuil $\alpha = 10\%$) de l'édition 2009 faisant écho aux sur-votes réciproques et aux blocs identifiés sur la période 1993-2008 (ici en grisé – cf Figure 15)

Parmi les 65 votes anormalement élevés de l'édition 2009, moins de la moitié (31) sont étrangers aux logiques de blocs précédemment identifiés. Ces votes, principalement occasionnels (moins du tiers d'entre eux confirment des sur-votes identifiés sur la période d'étude antérieure), peuvent être répartis en quatre grands groupes dont nous donnons le détail dans le Tableau 5.

Tableau 5 : votes anormalement élevés de l'édition 2009 externes aux blocs identifiés sur la période 1993-2008

votes entre pays voisins et/ou entre pays d'anciennes formations (URSS ou Yougoslavie)	Estonie → Russie Arménie → Russie Azerbaïdjan → Ukraine Pologne → Ukraine Hongrie → Ukraine Serbie → Bosnie-Herzégovine Bulgaire → Grèce Andorre → Espagne
votes de « rattachement », c'est-à-dire votes tentant d'intégrer des pays à de petites structures préexistantes	<i>rattachement de l'Allemagne au groupe de pays scandinaves et anglo-saxons :</i> Royaume-Uni → Allemagne Danemark → Allemagne Norvège → Allemagne Allemagne → Malte <i>rattachement du Monténégro au groupe des Balkans :</i> Monténégro → Croatie Monténégro → Bosnie-Herzégovine Monténégro → Albanie <i>rattachement de la Suisse au groupe des pays méditerranéens</i> Suisse → Espagne Suisse → Portugal Suisse → Albanie
votes communautaires en faveur d'Israël	Belgique → Israël France → Israël Andorre → Israël Bosnie-Herzégovine → Israël
votes « inclassables »	Islande → Portugal Bulgarie → Finlande Serbie → Malte République Tchèque → Arménie Slovaquie → Estonie Russie → France France → Portugal Slovénie → Danemark République Tchèque → Portugal

Bien que non directement concernés par les blocs, ces votes confirment certains des critères identifiés comme favorisant l'attribution de points, à savoir le voisinage géographique, l'appartenance à d'anciennes formations ou encore les votes communautaires (ici en faveur d'Israël – ce vote communautaire a déjà été identifié par (Gatherer, 2006) pour des éditions antérieures).

D'une manière générale, nous retiendrons que les votes anormalement élevés de l'édition 2009 :

- reflètent une tendance des pays à intégrer dans leur choix des critères autres que la qualité des chansons (plus de 15% des votes se trouvent au-delà du 9^e décile de leurs distributions théoriques) ;
- confirment le découpage de l'Europe en trois grands blocs ainsi que leurs extensions ;
- révèlent des patterns de votes occasionnels, pouvant être expliqués dans la plupart des cas par des critères classiques (dont le critère de voisinage).

Conclusion

L'étude des votes de l'ESC est un sujet intéressant à double-titre, car il confronte un réseau social (le graphe des votes entre les pays) et un réseau spatial (le graphe de voisinage des pays). Dans une approche *exclusivement quantitative*, nous avons proposé dans cet article de comprendre comment ces deux réseaux interagissaient en articulant notre travail en deux temps :

- dans un premier temps, nous avons mis en œuvre une méthode d'analyse statistique des votes permettant d'identifier les patterns de votes anormaux (en particulier le graphe des sur-votes sur la période d'étude 1993-2008) ;
- dans un second temps, nous avons démontré, grâce à un test statistique adéquat, que les sur-votes étaient significativement associés à des couples de pays géographiquement proches.

Cette construction est très ouverte, car elle nous a permis par la suite de mettre en évidence la manière dont ces sur-votes s'organisaient, tant structurellement que spatialement. À ce titre :

- nous avons identifié trois blocs de pays ayant une tendance significative à échanger des points (le bloc scandinave, le bloc des pays de l'ex-Yougoslavie associés à la Turquie et le bloc d'Europe orientale associé à Chypre et à la Grèce) ;
- nous avons confronté les votes de l'édition 2009 et confirmé les structures précédemment identifiées ainsi que les critères sous-tendant ces structures.

Au-delà de la réponse que cette étude apporte sur le lien longtemps pressenti entre « votes de faveur » et voisinage géographique à l'ESC ainsi que sur les alliances qui en découlent, elle propose une méthodologie originale permettant d'évaluer en quoi la spatialisation des individus intervient dans le réseau de relations qu'ils entretiennent. Ce type d'étude pourrait ainsi être appliqué à l'analyse des votes politiques au sein d'assemblées internationales (parlement européen, ONU) ou encore à l'analyse de l'attribution des points par des jurys internationaux lors de compétitions sportives.

Bibliographie

- Berge C., 1973. *Graphes*. Éditions Gauthier-Villars, Paris, 400 p.
- Dekker A., 2008. The Eurovision Song Contest as a 'Friendship' Network. *Connections*, Volume 29-I, 14 p.
- Fenn D., Suleman O., Efstathiou J. & Johnson N.F., 2005. How does Europe make its mind up? Connections, cliques and compatibility between countries in the Eurovision Song Contest. http://arxiv.org/PS_cache/physics/pdf/0505/0505071.pdf
- Freeman, 1979, Centrality in Social Networks – Conceptual Clarification. *Social Networks*, No.1, 215-239
- Gatherer D., 2006. Comparison of Eurovision Song Contest Simulation with Actual Results Reveals Shifting Patterns of Collusive Voting Alliances. *Journal of Artificial Societies and Social Simulation*, Volume 9, n°2
- Ginsburgh V. & Noury A., 2004. Cultural Voting. The Eurovision Song Contest. http://www.core.ucl.ac.be/services/psfiles/dp05/dp2005_6.pdf
- Good P., 1994. *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. Springer Verlag, New York, 228 p.
- Wasserman S. & Faust K., 1994. *Social Network Analysis – Methods and applications*. Cambridge University Press, 857 p.

- Wikipédia (contributeurs de), 2009a. Concours Eurovision de la chanson. *Wikipédia, l'encyclopédie libre*, page consultée le 20 août 2009
<http://fr.wikipedia.org/w/index.php?title=Concours_Eurovision_de_la_chanson&oldid=43697822>
- Wikipédia (contributeurs de), 2009b. Eurovision Song Contest. *Wikipédia, l'encyclopédie libre*, page consultée le 20 août 2009
<http://en.wikipedia.org/wiki/Eurovision_Song_Contest>
- Yair G., 1995. 'Unite Unite Europe' The political and cultural structures of Europe as reflected in the Eurovision Song Contest. *Social Networks* 17: 147-161

Annexe – Calcul des distributions théoriques des scores S_j et des votes V_{ij}

Dans cette annexe, nous travaillons dans le cadre d'une édition k donnée. Par souci de simplification, nous n'indiquerons pas les grandeurs utilisées dans ces calculs par l'édition k . Néanmoins, nous soulignons que ces calculs dépendent du nombre de pays votants et du nombre de pays compétiteurs, et donc de l'édition considérée.

Nous illustrerons et commenterons nos calculs en prenant pour exemple l'édition 2005 ayant rassemblé 39 pays – dont 24 compétiteurs.

Notations

Nous notons n le nombre de pays votants et m le nombre de pays compétiteurs de l'édition considérée. Jusqu'en 2003, ces deux nombres sont égaux ($n = m$), mais après 2003 et l'introduction des demi-finales, il existe des pays votants qui ne sont pas compétiteurs ($n > m$).

V_{ij} est la grandeur aléatoire représentative du vote que i attribue à j . Elle peut prendre les valeurs $v \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 10, 12\}$.

S_j est le score final obtenu par le pays j . Ce score est l'addition des votes émis par tous les pays votants, c'est-à-dire :

- les $(m-1)$ pays compétiteurs autres que j (j ne pouvant voter pour lui-même),
- les $(n-m)$ pays votants-non compétiteurs.

La distribution de la variable S_j nécessite d'être calculée par récurrence sur le nombre de pays contribuant au score final de S_j .

Nous notons donc par $S_j(x, y)$ le variable aléatoire représentative du score obtenu par j en cumulant les votes des x pays compétiteurs et les votes des y pays votants-non compétiteurs contribuant à ce score. En particulier, avec cette notation, le score final S_j de j s'écrit : $S_j = S_j(m-1, n-m)$.

Calcul de la distribution théorique du vote V_{ij} sans condition sur le score final S_j

• Premier cas de figure : le pays votant i est aussi compétiteur

Dans ce cas, i répartit dix votes non nuls ($v \in \{1, 2, 3, 4, 5, 6, 7, 8, 10, 12\}$) parmi $m-1$ pays compétiteurs (autres que i , car i ne peut voter pour lui-même) – par conséquent :

- $P(V_{ij [i \text{ compétiteur}] = v}) = \frac{1}{m-1}$ si $v \in \{1, 2, 3, 4, 5, 6, 7, 8, 10, 12\}$;
- $P(V_{ij [i \text{ compétiteur}] = v}) = \frac{m-11}{m-1}$ si $v = 0$.

- **Second cas de figure : le pays votant i n'est pas compétiteur**

Dans ce cas, i répartit dix votes non nuls ($v \in \{1, 2, 3, 4, 5, 6, 7, 8, 10, 12\}$) parmi les m pays compétiteurs – par conséquent :

- $P(V_{ij} [i \text{ non compétiteur}] = v) = \frac{1}{m}$ si $v \in \{1, 2, 3, 4, 5, 6, 7, 8, 10, 12\}$;
- $P(V_{ij} [i \text{ non compétiteur}] = v) = \frac{m-10}{m}$ si $v = 0$.

Nota : dans la suite, si le contexte est explicite, nous ne précisons pas la qualité de i (compétiteur ou non compétiteur) dans l'expression des probabilités de type $P(V_{ij} = v)$.

La Figure 17 présente la distribution théorique des votes V_{ij} pour l'édition 2005, dans l'hypothèse où l'attribution des points est complètement aléatoire.

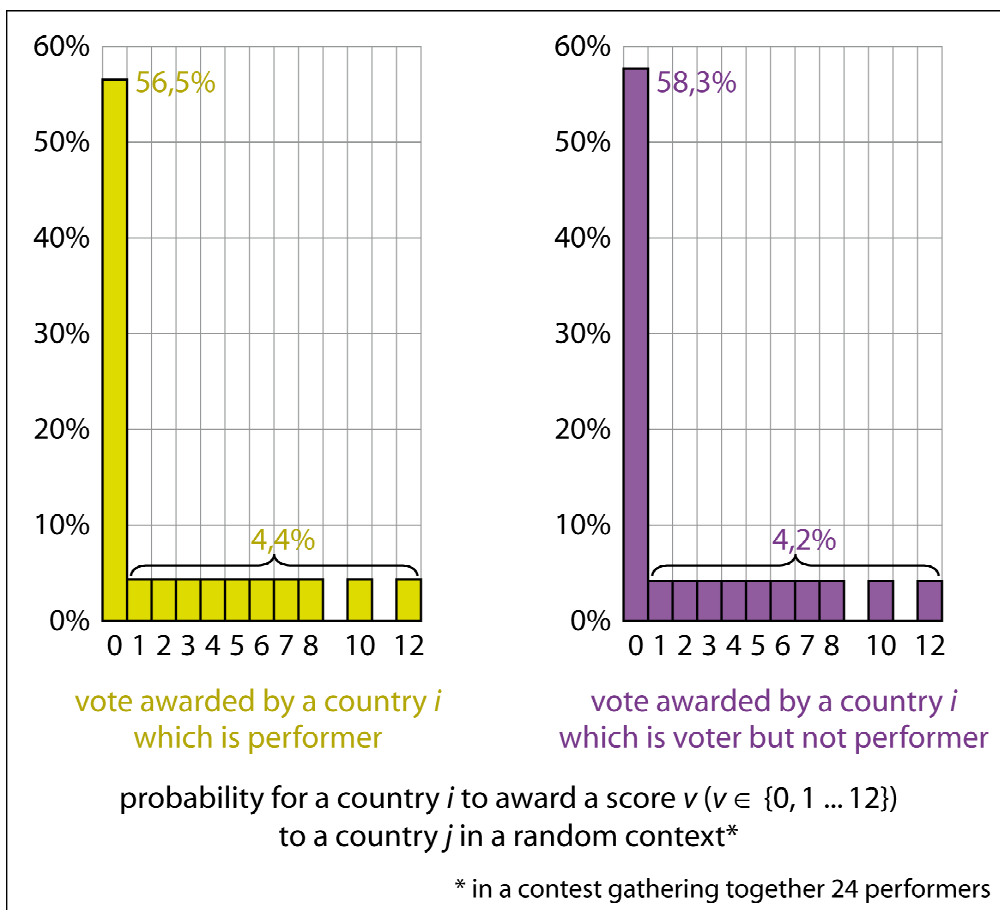


Figure 17 : distribution théorique du vote V_{ij} attribué par un pays i à un pays j dans l'hypothèse où l'attribution des points est aléatoire, pour une édition rassemblant 24 pays compétiteurs (type 2005) – cette distribution diffère légèrement selon que le pays i est compétiteur (graphique de gauche) ou non (graphique de droite)

Par construction, tous les votes non nuls ont une probabilité égale d'être attribués, quel que soit le score final obtenu par le pays destinataire (et donc quelle que soit la qualité perçue de la chanson).

Calcul de la distribution théorique du score final S_j

L'ensemble des pays votants-non compétiteurs et l'ensemble des pays compétiteurs étant disjoints, on peut écrire, grâce à la formule des probabilités totales, que :

$$P(S_j(x, y) = s) = \sum_{t=0}^s P(S_j(x, 0) = s - t) \times P(S_j(0, y) = t)$$

En d'autres termes : le score total s obtenu par j est la somme des contributions des x pays compétiteurs (à hauteur de $s-t$) et des contributions des y pays votants-non compétiteurs (à hauteur de t) pour toute valeur de t comprise entre 0 et s .

• Calcul de $P(S_j(x, 0) = u)$

La valeur de cette probabilité est déterminée par récurrence sur le nombre x de pays compétiteurs contribuant au score de j .

- initialisation de la récurrence : $P(S_j(1, 0) = u) = P(V_{ij [i \text{ compétiteur}]} = u)$;

- formule de récurrence :

$$\forall z \leq x : P(S_j(z, 0) = u) = \sum_{v \in \{0 \dots 12\}} P(S_j(z-1, 0) = u-v) \times P(V_{ij [i \text{ compétiteur}]} = v)$$

grâce à la formule des probabilités totales.

• Calcul de $P(S_j(0, y) = u)$

La valeur de cette probabilité est déterminée par récurrence sur le nombre y de pays votants-non compétiteurs contribuant au score de j .

- initialisation de la récurrence : $P(S_j(0, 1) = u) = P(V_{ij [i \text{ non compétiteur}]} = u)$;

- formule de récurrence :

$$\forall z \leq y : P(S_j(0, z) = u) = \sum_{v \in \{0 \dots 12\}} P(S_j(0, z-1) = u-v) \times P(V_{ij [i \text{ non compétiteur}]} = v)$$

grâce à la formule des probabilités totales.

• Calcul de $P(S_j = s)$

Étant donné que le score final S_j correspond à la variable aléatoire $S_j(m-1, n-m)$, sa distribution théorique se calcule grâce à la formule de $P(S_j(x, y) = s)$ explicitée ci-dessus, avec $x = m-1$ et $y = n-m$.

La Figure 18 présente la distribution théorique du score final S_j obtenu par un pays j pour l'édition 2005 ($n = 39$, $m = 24$) en considérant que les pays attribuent leurs votes aléatoirement.

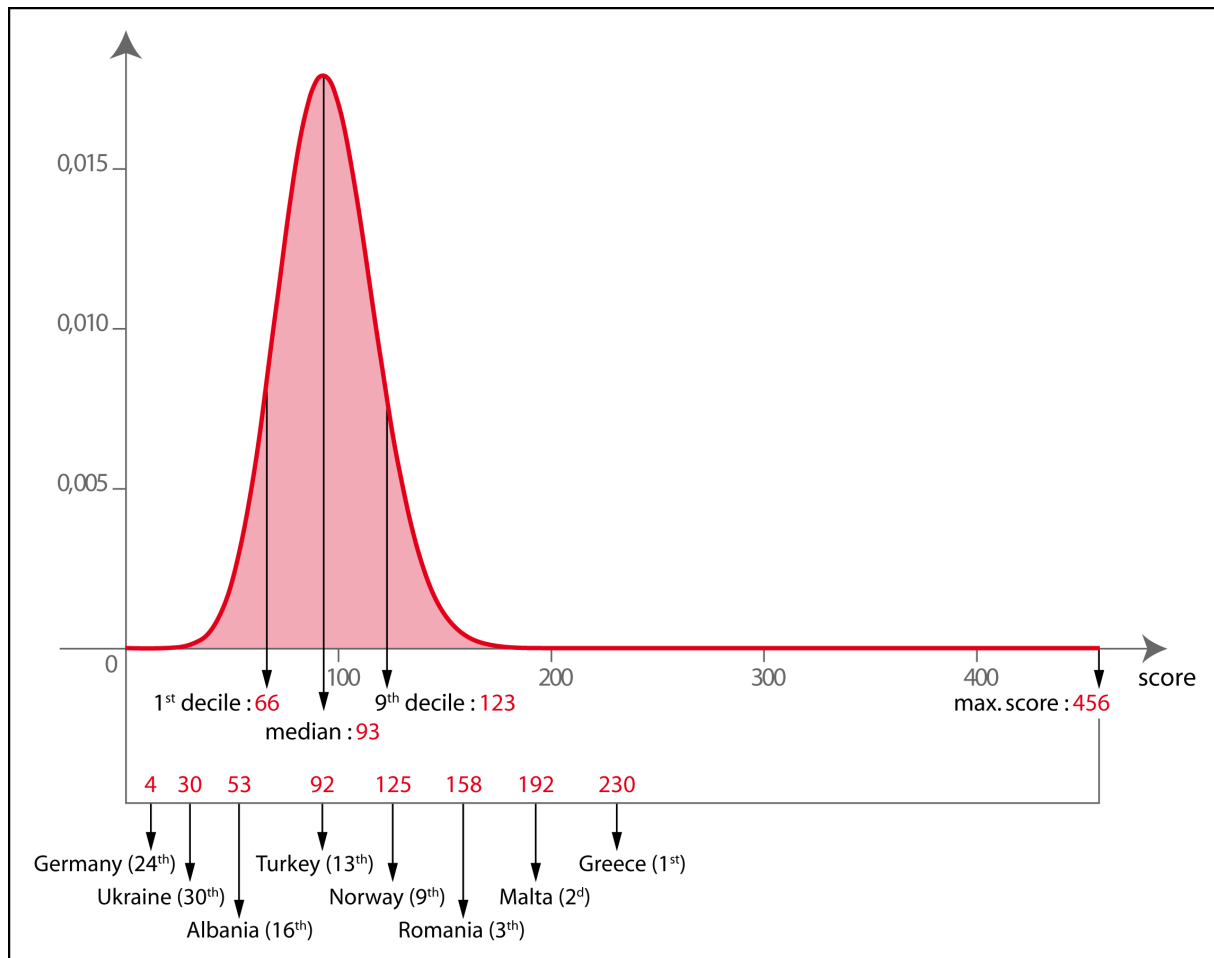


Figure 18 : distribution théorique du score final S_j obtenu par un pays compétiteur j lors d'une édition rassemblant 39 pays, dont 24 compétiteurs – ce format d'édition s'est produit en 2005 : la partie inférieure du graphique situe les scores et les classements de certains pays lors de cette édition

À noter que les scores relevés lors de l'édition 2005 sont beaucoup plus étalés que le laisse prédire la distribution théorique : les médianes théorique et observée sont à peu près égales, mais on dénombre dix compétiteurs obtenant un score final supérieur au 9^e décile et symétriquement dix compétiteurs obtenant un score final inférieur au 1^{er} décile.

En d'autres termes, seul un compétiteur sur six obtient un score final compris entre le 1^{er} et le 9^e déciles : ce constat (que l'on pourrait établir pour chacune des éditions) met en évidence que les distributions des scores obtenus par les compétiteurs ne découle pas d'une attribution aléatoire des points.

Calcul de la distribution théorique du vote V_{ij} conditionnellement au score final S_j

- **Premier cas de figure : $n = m$ (éditions antérieures à 2003 inclus)**

On cherche donc à calculer $P(V_{ij} = v \mid S_j(m-1, 0) = s)$, avec i pays compétiteur.

Grâce à la formule des probabilités conditionnelles, on obtient :

$$P(V_{ij} = v \mid S_j(n-1, 0) = s) = \frac{P[(V_{ij} = v) \cap (S_j(n-1, 0) = s)]}{P(S_j(n-1, 0) = s)}$$

L'événement $(V_{ij} = v) \cap (S_j(n-1, 0) = s)$ est équivalent à l'événement $(V_{ij} = v) \cap (S_j(n-2, 0) = s - v)$ car $S_j(n-1, 0) = S_j(n-2, 0) + V_{ij}$.

Les votes du jury i et des $n-2$ autres jurys étant indépendants, on a :

$$\begin{aligned} & P[(V_{ij} = v) \cap (S_j(n-1, 0) = s)] \\ &= P[(V_{ij} = v) \cap (S_j(n-2, 0) = s - v)] \\ &= P(V_{ij} = v) \times P(S_j(n-2, 0) = s - v) \end{aligned}$$

et donc :

$$P(V_{ij} = v \mid S_j(n-1, 0) = s) = \frac{P(V_{ij [i \text{ compétiteur}] = v}) \times P(S_j(n-2, 0) = s - v)}{P(S_j(n-1, 0) = s)}$$

• **Second cas de figure : $n > m$ (éditions postérieures à 2003)**

On cherche donc à calculer $P(V_{ij} = v \mid S_j(m-1, n-m) = s)$, avec deux sous cas :

- i pays compétiteur,
- ou i pays non compétiteur.

En effectuant un raisonnement analogue à celui mené précédemment, on obtient grâce à la formule des probabilités conditionnelles :

- si i est un pays compétiteur :

$$P(V_{ij} = v \mid S_j(m-1, n-m) = s) = \frac{P(V_{ij [i \text{ compétiteur}] = v}) \times P(S_j(m-2, n-m) = s - v)}{P(S_j(m-1, n-m) = s)}$$

- si i est un pays non compétiteur :

$$P(V_{ij} = v \mid S_j(m-1, n-m) = s) = \frac{P(V_{ij [i \text{ non compétiteur}] = v}) \times P(S_j(m-1, n-m-1) = s - v)}{P(S_j(m-1, n-m) = s)}$$

Les membres de ces expressions sont tous déterminés par les calculs sur la distribution théorique de V_{ij} sans condition sur S_j et par les calculs sur la distribution théorique de S_j .

La Figure 19 présente la distribution théorique des votes V_{ij} pour l'édition 2005 ($n = 39, m = 24$) conditionnellement au fait que le score final S_j obtenu par le pays j soit égal à 192 (cas de Malte pour cette édition).

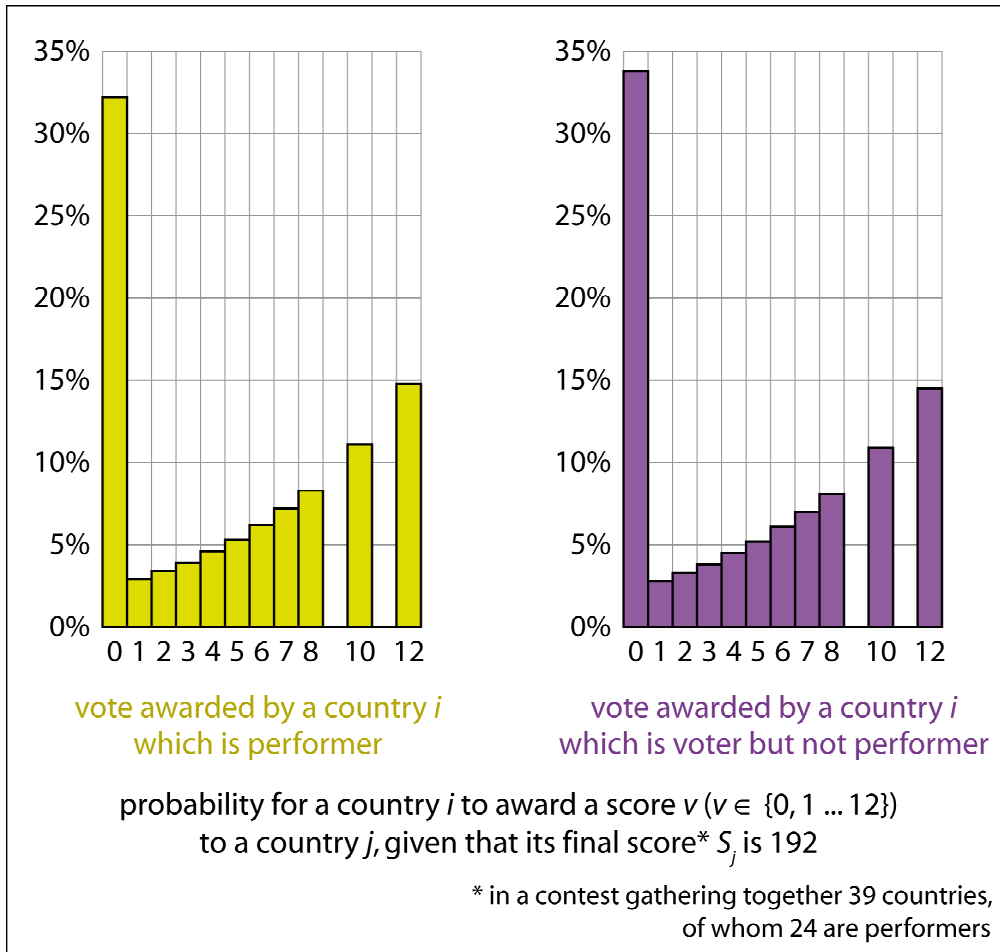


Figure 19 : distribution théorique du vote V_{ij} attribué par un pays i à un pays j sachant que ce dernier obtient un score total S_j de 192, pour une édition rassemblant 39 pays, dont 24 compétiteurs (type 2005) – cette distribution diffère selon que le pays i est compétiteur (graphique de gauche) ou non (graphique de droite)

En tenant compte de la qualité perçue de la chanson (au travers du score final obtenu par le pays destinataire du vote), les votes les plus probables pour un pays recueillant au final 192 points sont les deux votes les plus élevés (12 et 10 points) et le vote nul.

Les fortes probabilités associées aux votes élevés s'expliquent par l'importance du score 192 (dont nous avons vu en Figure 18 qu'il était bien au-delà du 9^e décile de la distribution de S_j dans l'hypothèse de votes aléatoires).

Par ailleurs, la forte probabilité associée au vote nul est caractéristique de la règle d'attribution des votes, qui exige de chaque pays de n'attribuer qu'une dizaine de votes non nuls (et donc des votes nuls à tous les autres pays).

La prise en compte du score final S_j dans le calcul de la distribution de V_{ij} permet d'apprécier la vraisemblance du vote que i accorde à j . Dans cet exemple, compte-tenu du score final élevé obtenu par j :

- les votes faibles ($v = 1, 2$ ou 3) apparaissent peu vraisemblables, car peu récurrents dans les combinaisons de 39 votes permettant d'atteindre le score 192 ;
- et inversement pour les scores élevés ($v = 10$ ou 12) et le score nul ($v = 0$).