

Measuring the Broadband Access Divide

Michael Paul Fourman
School of Informatics
10 Crichton Street
Edinburgh, Scotland
michael.fourman@ed.ac.uk

ABSTRACT

Decreases in a Gini index for broadband uptake have been interpreted as evidence of a narrowing digital divide. Nevertheless, a significant divide persists.

How should we measure the divide? We propose two related indices, introduced in the context of health inequality by Wagstaff *et al.* (1991, 2005), as measures for the depth and breadth of the digital divide. We show how these quantify the contribution of the digital divide to social inequalities and cycles of deprivation. Depth measures the barriers to digital inclusion presented by existing deprivation. Breadth measures the degree to which the digital divide tends to reinforce existing inequalities.

We report briefly on two applications, one local, one global, to illustrate how these measures can be used to assess progress and inform policies intended to reduce the digital divide.

CCS Concepts

•Social and professional topics → Broadband access;

Keywords

Gini, digital, divide, global, local, inequality, deprivation.

1. INTRODUCTION

In his millennial State of the Union address, President Clinton announced tax incentives intended, *to close the digital divide and open opportunity for our people.*

Opportunity for all requires something else today—having access to a computer and knowing how to use it. This means that we must close the digital divide...
Bill Clinton, 2000 [3]

In 2002, a report from the U.S. Department of Commerce (DoC [8]) adapted the standard methodology for assessing the distribution of income, to produce a *Gini Coefficient for Computer and Internet Use*. This adaptation is identical to the **concentration index**, C , of Wagstaff *et al.* 1991 [10].

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

ICTD '16 June 03-06, 2016, Ann Arbor, MI, USA

© 2016 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-4306-0/16/06.

DOI: <http://dx.doi.org/10.1145/2909609.2909616>

Cumulative benefit (population online) is plotted against cumulative population, ordered by income, to give a *Lorenz Curve*. Fig. 1 shows an example, using data for computer use in 1997 (DoC *op. cit.* p.24 Fig 2-1).

Inequality is indicated by deviation from the equally-dashed diagonal line representing perfect equality, and measured as the difference between the areas above and below the Lorenz curve.¹ This difference is divided by the area, p , of the enclosing rectangle² to give a concentration coefficient of 19%.

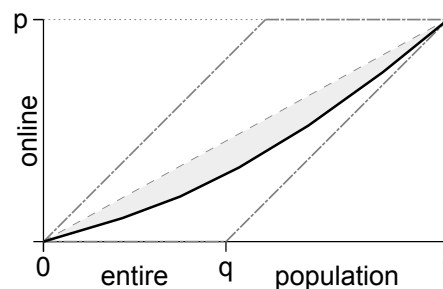


Figure 1: Concentration Index

The DoC report paints a rosy picture. Decreasing values of the index are interpreted to conclude, *inter alia*, that from 1984 to 2001 *the distribution of computers among households has moved continuously in the direction of less inequality.*

Others, e.g. Sciadas and Cho[6, 2], have used similar methods to reach similar conclusions. However, Sciadas comments that, *The lowest income groups... continue to lose ground vis-à-vis the very high income groups*, whereas his computation of the Gini index, for the distribution of computer use against income, suggested that the digital divide was generally closing. Kelly [4], commenting on Cho's finding of a reduction in the global divide, says, *other evidence suggests that the progress in reducing the digital divide has occurred mainly as a result of middle-income countries catching up, whereas some of the least developed countries have actually been falling behind.*

The Gini index failed to capture perceived increases in the digital divide. Nevertheless, a 2009 United Nations report again used dramatic reductions in a Gini index — which

¹This difference can be seen to be twice the shaded area between the Lorenz curve and the line of perfect equality.

²Here, p is the proportion of the population online; $q = 1 - p$ the proportion offline. The parallelogram is for later use.

dropped from 95% in 2000 to less than 50% in 2008—to justify a claim that, *Inequality is shrinking*. ([7] p. 16)

A 2015 report from UNESCO and the ITU [1] says, *the digital divide is proving stubbornly persistent in terms of access to broadband Internet*. A 2016 report from the World Bank [5] says, *digital divides persist across income, age, geography, and gender*. While these more-recent publications still recognise a persistent divide, neither mentions the Gini index—nor do they suggest other measures for the divide.

Renormalisation.

In 2005, Wagstaff [9] observed that, when the advantage considered is binary—as broadband uptake is—the concentration index, C , a relative measure of inequality, must be renormalised. Wagstaff *et al.* 1991 [10] also introduced a generalised concentration index (GCI) as an absolute measure of inequality. This must also be modified—in this case, simply scaled—for application to a binary advantage.

We will call the renormalised concentration index, D , the **depth** of the divide, and the scaled GCI, B , its **breadth**—we explain these names below. Each has a simple algebraic definition, in terms of C , the concentration index, and p , the proportion of the population enjoying the advantage. If $q = 1 - p$ is the proportion excluded, then,

$$D = C/q \quad B = 4Cp = 4Dpq \quad (1)$$

Fig. 2 shows data from the DoC report— p is home computer uptake (p.3 Fig. 1-1); C is the concentration index (p.87 Fig. 9-3)—together with our computed values for B and D . From 1990 to 2002, the concentration index, C , for

Y	1990	1994	1998	1999	2001	2002
p	15.9%	22.6%	36.6%	42.1%	51.0%	56.5%
C	40%	39%	31%	30%	26%	23%
B	25.4%	35.3%	45.4%	50.5%	53.0%	52.0%
D	47.6%	50.4%	48.9%	51.8%	53.1%	52.9%

Figure 2: Home computer uptake (from DoC data)

Households with a Computer, plotted against Income, fell consistently. Breadth and depth tell a different story.

Outline.

In §2 we show that both breadth and depth arise as natural measures of the effects of the divide on inequality. They can also be used to identify the places where we must increase uptake, in order to close the divide.

In §3.1 we use postcode-level data for Scotland to relate digital exclusion to the Scottish Index of Multiple Deprivation (SIMD). In §3.2 we apply these indices to ITU data, and discuss their interpretation in that context.

2. QUANTIFYING THE DIGITAL DIVIDE

Digital inclusion affords increased opportunities, in health, education, social inclusion, and well-being, to individuals in all sectors of society. However, many factors of deprivation constitute barriers to digital inclusion. So the benefits of increasing inclusion often serve to widen the opportunity gap, and to reinforce existing inequalities. To assess the social impacts of the digital divide, we quantify these effects.

Abstractly, we consider the effects of some binary advantage on the relationships between individuals from a popu-

lation subject to some deprivation ordering, $\prec$, where $b \prec a$, (b is below, and a above), if b is more deprived than a .

Concretely, our individuals are households who may be online or offline. For each offline-online pair, (u, v) , of households, if the offline household is inferior ($u \prec v$), then the digital gap between these two households strengthens v 's superiority. On the other hand, if $v \prec u$, then v 's digital advantage provides opportunities that serve to reduce the existing inferiority.

We divide the set of all offline-online pairs into S , those that strengthen deprivation, and R , those that reduce it.

$$\begin{aligned} S &= \{(u, v) \mid u \text{ is offline, } v \text{ is online, } u \prec v\} \\ R &= \{(u, v) \mid u \text{ is offline, } v \text{ is online, } v \prec u\} \end{aligned} \quad (2)$$

If the distribution of broadband uptake were independent of deprivation, we should expect these two sets to have the same size. In general, wherever the dependence of uptake on deprivation has been studied, S is larger than R .

The excess of S over R provides a natural measure of deprivation dependence. To give a normalised index that is independent of the size of the population, we divide $(S - R)$ by the number of possible pairs, then scale to give an index that occupies the range $[-1, 1]$. Our two indices are defined by entertaining two different sets of possibilities. If $\mathbf{N}$ is the total number of individual households, we define,

$$D = \frac{S - R}{S + R} = \frac{S - R}{pq\mathbf{N}^2} \quad B = \frac{S - R}{\mathbf{N}^2/4} \quad (3)$$

The depth index considers only the offline-online pairs. The breadth index considers all pairs of households.

We will now show that these are precisely Wagstaff's indices (1). Consider again Fig. 1. The pecked lines along the top and bottom of the parallelogram represent the Lorenz curves for two extremely unequal distributions.³ In one extreme, represented by the lower line, each offline household would be more deprived than every online household. This Lorenz curve follows the horizontal axis through the offline population (of size $q = 1 - p$), and then rises, with slope 1, through the online population. The curve for the other extreme, in which the most deprived sections of the population are online, traces the top of the parallelogram. The coefficient C has range $[-q, q]$. Wagstaff proposed the renormalisation $D = C/q$, to give D the range $[-1, 1]$.

This amounts to dividing the difference in areas above and below the Lorenz curve by the area of the parallelogram of Fig. 1, instead of the area of the rectangle. For a point (x, y) on the Lorenz curve x is cumulative population, and y cumulative online population. We transform the parallelogram to a rectangle, and represent the same curve on a plot of cumulative online population, $v = y$, against cumulative offline population, $u = x - y$.

This is shown on the left-hand side of Fig. 3. The rectangle here represents the set of offline-online pairs, (u, v) , of households, sorted in each dimension by our deprivation ordering, $\prec$. The Lorenz curve separates the pairs in R , above the curve, with $u \prec v$, from those in S , below, with $v \prec u$.

From this presentation, it is straightforward to compute that the **depth** index represents the expected level of deprivation of an offline household, relative to the population

³The values $p = 0.54$ and $q = 0.46$ represent the proportions of the population online and offline. In September 1997, 54% of the US population aged 3 and over were computer users.

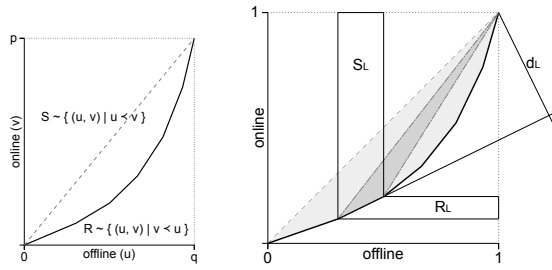


Figure 3: Depth of the Divide

of online households, on a $[-1, 1]$ scale (1 is most deprived). In so far as the various factors of deprivation act as barriers to digital inclusion, this provides a measure of the obstacles that must be overcome to get each offline individual online.

The **breadth** of the divide is an absolute measure of the degree to which the digital divide acts to strengthen existing divides. The generalised concentration index compares $S - R$ with the total number of pairs N^2 . Thus it measures the net effect of the digital divide on all possible binary interactions.

An extreme case for the breadth index occurs when the more deprived half of the population is offline and the less deprived is online (or vice-versa). We have $N^2/4$ offline-online pairs, and they all fall in the set S where digital disadvantage acts to strengthen (or reduce) existing deprivation. Thus, if we apply Wagstaff's generalised concentration index to a binary advantage, the factor of 4 is required to give an index on a $[-1, 1]$ scale.

The right-hand graph of Fig. 3 shows again the same curve scaled to the unit square. This has all the advantages of the traditional Gini plot. The depth index is represented as twice the area between the Lorenz curve and the line of perfect equality, and we can plot and compare curves for different populations and different years on the same diagram.

We use this diagram to quantify the distribution of inequality. The curve is made up of line segments, L , each representing a segment of the population. The net weight of digital disadvantage on one segment of the population is represented by a difference: the area above it, S_L , representing the pairs in S whose offline member is in L , minus the area to its right, R_L , representing the pairs in R whose online member is in L . The area of the shaded triangle is half of this difference. The depth of L 's digital disadvantage is represented by d_L , the height of the triangle.

We can also use this analysis to focus efforts to close the divide. If L is some segment of a large population, defined by its deprivation ranking, and there are more online households above L than offline households below L , then the marginal effect of moving an offline household in L online will be to decrease the breadth of the divide. Similarly, we can compute a threshold for the difference between the numbers of households online above, and offline below L , above which the marginal effect of moving an offline household in L online will be to reduce the depth of the divide.

2.1 Related work

Both breadth and depth are closely related to the Gini index. The (similarly-related) decile dispersion, Palma, and 20/20 ratios each tell us something about the two ends of the Lorenz curve, but they ignore the middle ground.

Several other well-known indices are not relevant to our

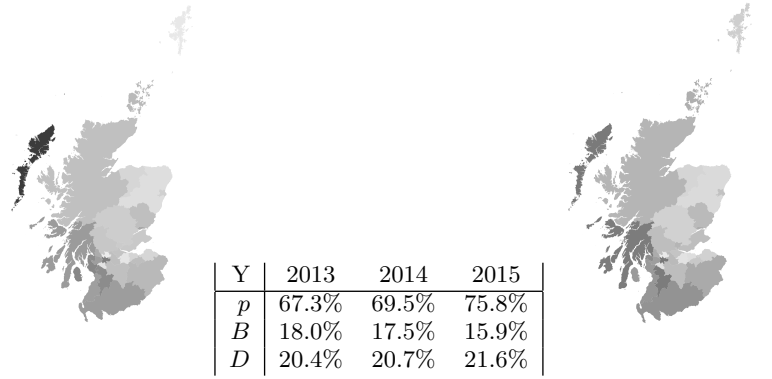


Figure 4: Households connected in Scotland

interest in the effects on inequality of the distribution of a potentially universal binary advantage. The Atkinson and Hoover indices both relate to (re-)distributable exclusive goods. The Theil index and general entropy measures capture variation at different scales, but ignore deprivation.

3. EXAMPLES

We have already seen that the simple definitions (1) of B and D make it easy to compute values for breadth and depth from the results of earlier studies.

For the DoC data in Fig. 2, we interpret the increase in depth of the divide to indicate that those who remained offline were increasingly those who faced the highest barriers. The growing breadth of the divide indicates an increasing national impact of digital disadvantage.

We briefly describe two examples, to indicate how an analysis of primary data can yield more information.

3.1 Scotland's Divide

Detailed data on broadband connections is recorded by service providers, for their own business purposes. In the UK, Ofcom has recently started to publish data giving the number of broadband connections in each postcode. Each of the five national ISPs, who together cover 90% of domestic connections, provides data which is then aggregated by Ofcom. We have combined this at postcode-level with data for the Scottish Index of Multiple Deprivation (SIMD), and census data (giving numbers of households). The data set covers around 180K postcodes, including a total of 2.5M households, over three successive years, 2013–2015.⁴ Fig. 4 shows that uptake has risen; the divide has become narrower, but deeper.

The maps show the relative depth of disadvantage (darker is deeper) experienced in each of Scotland's 32 local authority areas, in 2013 on the left, and 2015 on the right. We see that the position of the Western Isles, relative to rest of Scotland, has improved. However, East Renfrewshire, Glasgow City, Argyll and Bute, and the Western Isles, still suffer disproportionate shares of Scotland's digital disadvantage.

3.2 The Global Divide

To compute breadth and depth we require data on numbers of connections and numbers of households. We use ITU figures for broadband uptake. Numbers of households are computed, by division, from World Bank Total Population

⁴This work contains public sector information licensed under the UK Open Government Licence v2.0.

data,⁵ combined with household size data for 68 countries from 2000–2012 assembled and interpolated by TekCarta,⁶ which we have extrapolated to 2013/14. We order these countries by level of broadband uptake (per household) and plot cumulative proportion of online households against cumulative proportion of offline households.

Fig. 5 shows % figures for depth, D , breadth, B , and uptake, p , for households connected, across the 68 countries covered by the data, for the years 2000–2014. We see that the depth of the divide reduced annually in the period 2000–2011, but has increased since then. Meanwhile, the breadth of the divide has steadily increased.

These figures provide a lower bound for the global digital divide. They ignore within-country inequalities, and many poorly-connected, countries for which we have no data.

The Lorenz curve for each year includes a line segment for each country. If moving more people in that country online would increase the divide, this is indicated by the stroke: grey for increasing breadth; dotted for increasing depth.

The statistics for five selected years are highlighted in bold in the table. Each successive curve from 2001 to 2004 dominates the 2000 curve and its predecessors. The solid portions of the corresponding curves have thicker strokes for successive years. The curve for 2000 shows the greatest depth; the curve for 2011 shows the least. The curves for 2005–2007 show a decrease in D , and a changing pattern of inequality, with no Lorenz domination. From 2007–2011 we again see successive curves that dominate their predecessors. Finally, from 2011 to 2014 the depth of the divide increases, as more people go online in the more connected countries.

The Lorenz curves illustrate precisely the phenomenon remarked on by Kelly, the middle group of countries, is catching up, while the least developed countries are falling behind. The 2014 curve is approximated by three straight line segments, corresponding to three groups of countries. Their different slopes show different levels of opportunity.

Roughly 50% of the online households are in a group of well-connected countries that includes the USA and much of Europe. It accounts for only 10% of the offline households. In these countries the odds of being online are over 3 : 1. The next 45% of the online households are found in a group of moderately-connected countries, dominated by China, that accounts for around 40% of those offline, but is fast catching up with the leaders. Their odds of being online are roughly 3 : 4. The final, poorly connected group, which includes India, for example, includes around 5% of the online households, and 45% of those offline. In one of these countries, your odds of being online are roughly 3 : 40.

The 2015 State of Broadband report [1] says, *Network effects and externalities that multiply the impacts of ICTs require minimum adoption thresholds before those impacts can begin to materialize*, and suggests that, *multiplier effects may be widening the overall digital divide at a greater rate than simple adoption numbers suggest*. Ordering countries by rate of uptake means that our depth measure captures the level of digital disadvantage affecting those still offline.

Our analysis shows that to reduce the breadth of the global divide it will be necessary to increase uptake in countries in the third group. However we interpret recent increases in the depth of the global divide as evidence that

The digital divide across 68 countries,
from ITU Fixed Broadband data, 2000 - 2014.

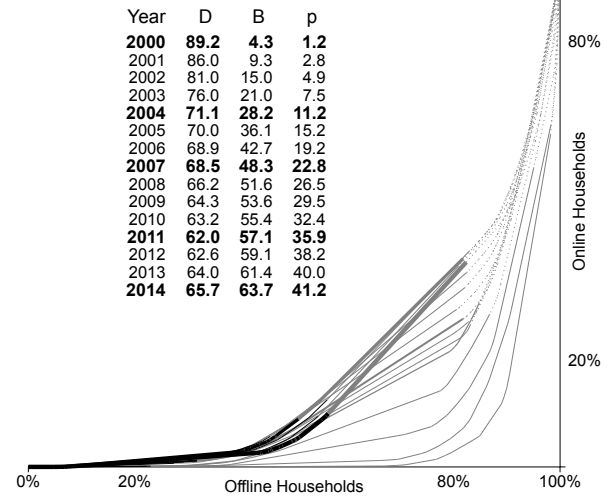


Figure 5: The Global Digital Divide 2000-2014

these countries are increasingly disadvantaged relative to those online. This will make it increasingly difficult for countries with little or no digital infrastructure to bootstrap their own digital inclusion.

4. CONCLUSION

Breadth and depth provide better measures than the Gini, of our still-faltering progress on digital inclusion.

5. REFERENCES

- [1] BROADBAND COMMISSION FOR DIGITAL DEVELOPMENT. The state of broadband, 2015.
- [2] CHO, C. How to measure the digital divide? ITU/KADO Symposium on Building Digital Bridges, September 2004.
- [3] CLINTON, W. J. Address before a joint session of the congress on the state of the union, January 2000.
- [4] KELLY, T. Twenty years of measuring the missing link. In *Maitland+20: Fixing the missing link*, G. Milward-Oliver, Ed. Anima Centre Limited, Bradford on Avon, 2005, pp. 23–33.
- [5] MISHRA, D., DEICHMANN, U., ET AL. World development report: digital dividends, 2016.
- [6] SCIADAS, G. Unveiling the digital divide. Statistics Canada, October 2002.
- [7] UN ICT ANALYSIS SECTION. Information economy report 2009: Trends and outlook in turbulent times. UN Conference on Trade and Development, 2009.
- [8] VICTORY, N. J., AND COOPER, K. B. A nation online: How Americans are expanding their use of the Internet, February 2002.
- [9] WAGSTAFF, A. The bounds of the concentration index when the variable of interest is binary, with an application to immunization inequality. *Health Economics* (2005), 429–432.
- [10] WAGSTAFF, A., PACI, P., AND VAN DOORSLAER, E. On the measurement of inequalities in health. *Social Science & Medicine* 33, 5 (1991), 545 – 557.

⁵<http://data.worldbank.org/indicator/SP.POP.TOTL>

⁶<http://www.generatorresearch.com/tekcarta/databank/households-average-household-size/>