PNAS
www.pnas.org

Supplementary Information for

The exploration of the chemical space and its three historical regimes

Eugenio J. Llanos, Wilmer Leal, Duc H. Luu, Jürgen Jost, Peter F. Stadler, and Guillermo Restrepo

Guillermo Restrepo.
E-mail: restrepo@mis.mpg.de

This PDF file includes:

Variance analysis of the number of new compounds

Table S1 contains the results of the Shapiro-Wilk and Kolmogorov-Smirnov normality tests for the growth rate r_t and for $Y_t = \ln s_{t+1} - \ln s_t$. These results show that r_t is not normally distributed (Figure S1a), but Y_t is for the three historical regimes (Table 1, main text; Table S1, and Figure S1). The distribution and its density of the combined residuals $Z_t := \frac{Y_t - \mu_i}{\sigma_i}$, where μ_i, σ_i are given in Table 1, are shown in Figure S2.

Table S1. Results of the Shapiro-Wilk and Kolmogorov-Smirnov normality tests for r_t, Y_t, and Z_t.

	r_t		Y_t		Z_t	
Period	p^{ShW}	p^{KS}	p^{ShW}	p^{KS}	p^{ShW}	p^{KS}
1804-2015	$< 2.2 \times 10^{-16}$	$< 5.54 \times 10^{-14}$	$< 4.143 \times 10^{-15}$	$< 1.013 \times 10^{-6}$	0.01934	0.2637
1804-1860	$< 6.986 \times 10^{-10}$	0.02414	0.05267	0.5133	0.05267	0.5133
1861-1980	0.001877	0.2138	0.0701	0.3391	0.07010	0.3391
1981-2015	0.2027	0.7196	0.08297	0.6242	0.08297	0.6242

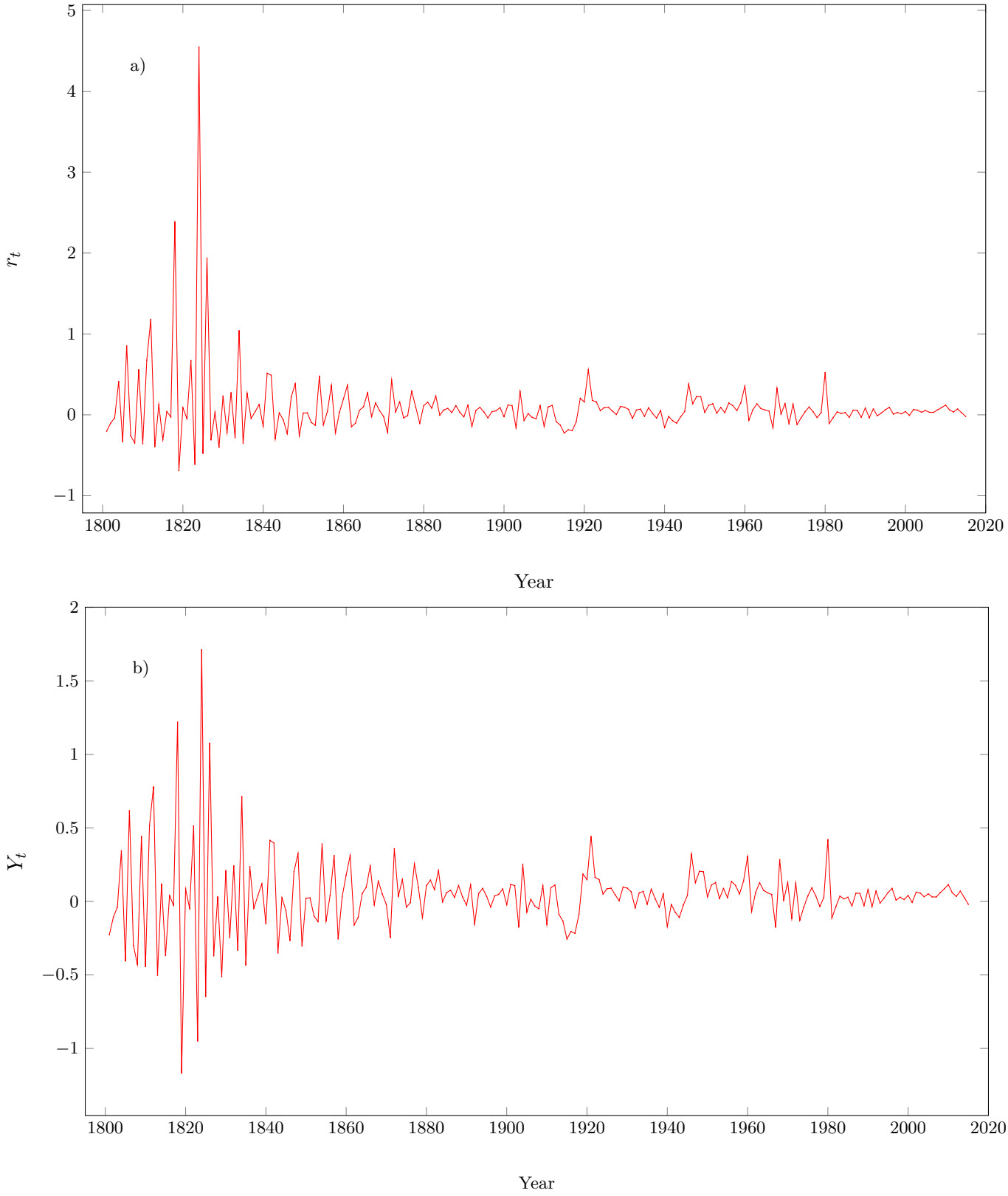

Fig. S1. a) Growth rate $r_t = (s_{t+1}/s_t) - 1$ and b) logarithm difference of number of new substances $Y_t := \ln s_{t+1} - \ln s_t$ for the period 1800-2015.

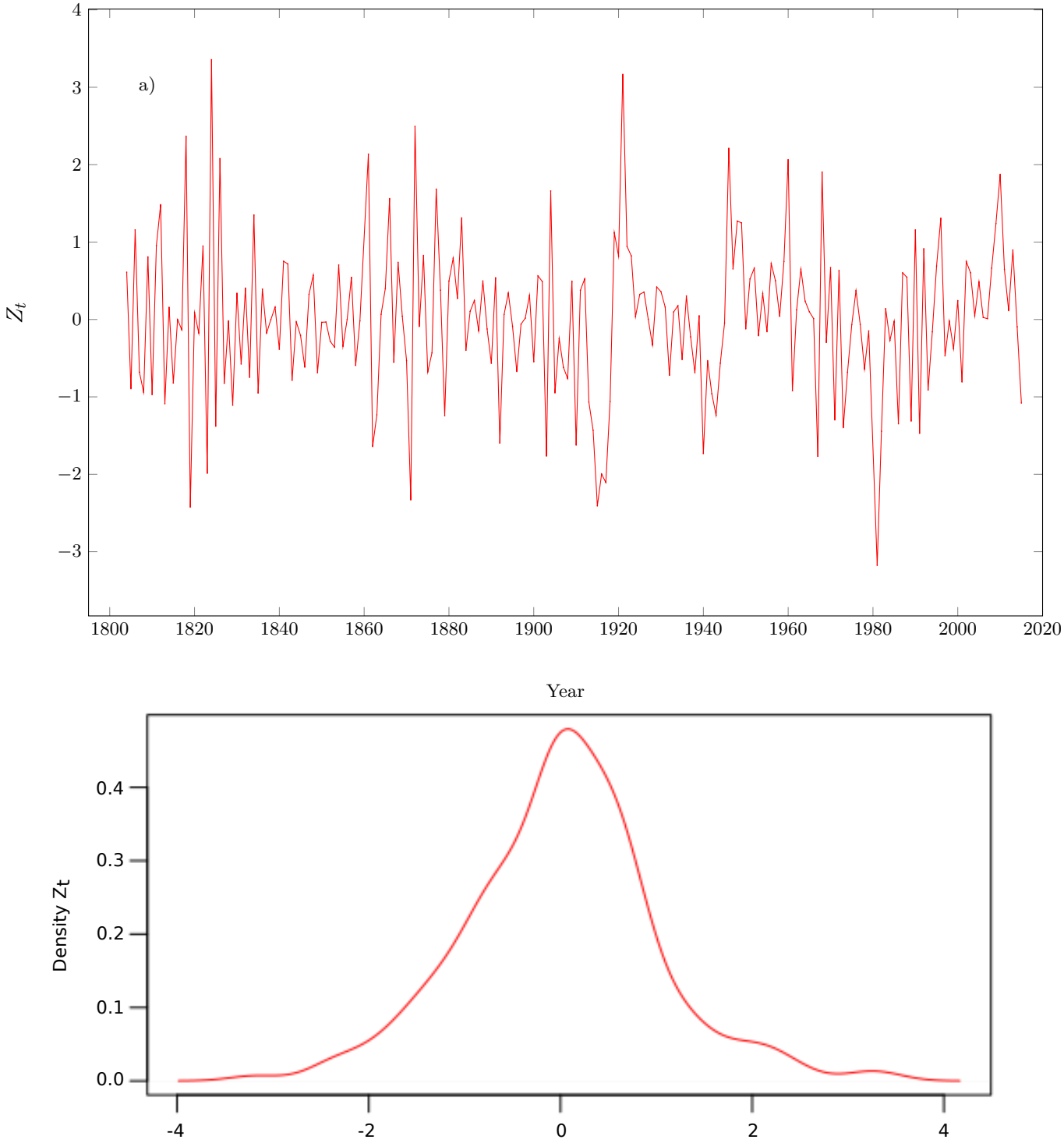

Fig. S2. a) The adjusted noise of logarithm difference during the period 1800-2015. b) The density of residuals of logarithm difference during the period 1800-2015.

Preferred compositions over time

For each compound reported for the first time in year t we extracted the combination of elements appearing in the formula of the compound and arranged the elements in lexicographic order, which we call composition. Hence, for HCl and H_2SO_4, the respective compositions are ClH and HOS. Table S2 shows the relative importance of inorganic compositions in the first regime, and Table S3 the dominance of organic compositions. Likewise, Table S4 shows the preponderance of organic chemistry compositions in the third regime, where all frequent metal compositions include C, highlighting the organometallic character of the third regime.

Table S2. List of the 10 most frequent compositions of elements in some particular years of the first regime. The second half of the table lists the remaining 10 most frequent compositions including metals. Non-carbon compositions in red.

1800	1810	1820	1830	1840	1850	1860
CHNaO	CHAlO	CHNO	CHNO	CHO	CHN	CHO
CHNO	CHNO	FeO	HOSZn	CHNO	CHNO	CHNO
CHCuO	CHKO	CFeN	HIKO	CHClNO	CHClNPt	CHN
OS	OPb	AuHO	ClHHgN	CHClO	AsCaHO	CHClO
CuHOS	CBaO	CBaN	OPZn	FeHKOS	HNOPt	CHOS
S	Pb	Fe	ClHIKO	CClO	CHFeNO	CHCl
OSn	KTe	FeS	CuOS	CHOS	CHO	IKSb
ClCuHO	CFeNS	CHFeNO	AsO	CHBrNO	CHOS	CHBrP
ClCu	NaTe	CrO	ClHO	CHCuNO	CHClN	IRbSb
NaOS	HTe	CCoO	CHO	NaOSb	CHNOPt	HINSb
CHgNO	CHNaO	HNaOS	CHNZn	CHKO	HNaOSe	ISbTl
ClHNaOPt	CHCuO	FeS	CHNiO	CHClOPt	BHOSr	CHOPb
-	CuHOS	Fe	HNaOP	ClHNOPt	AsBaHO	CuHOS
-	OSn	CaClHO	HNiO	ClHNPt	BHMgO	CHOSn
-	ClCuHO	AgOS	CuHOS	BHMgO	CFeNO	CHCaO
-	ClCu	P	FeHOS	HNaOP	CHCuNO	NaOSi
-	NaOS	MgO	CoNO	CHFeNO	ClHNOPt	CaHOSe
-	CHgNO	CKO	HNOZn	CHCuO	ClHNPt	CHSn
-	ClHNaOPt	CrS	ClCoHN	CHAgNOS	HNOPtS	CHISn
-	-	CHKMgO	BaHOP	CHOPb	CoS	HNaOS

Table S3. List of the 10 most frequent compositions of elements in some particular years of the second regime. The second half of the table lists the remaining 10 most frequent compositions including metals. Non-carbon compositions in red.

1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980
CHNO	CHO	CHNO	CHNO	CHNO	CHNO	CHNO	CHNO	CHNO	CHNO	CHNO	CHNO
CHO	CHNO	CHO	CHO	CHO	CHO	CHO	CHO	CHO	CHO	CHO	CHO
CHClO	CHNOS	CHN	CHNOS	CHNOS	CHN	CHNOS	CHNOS	CHNOS	CHNOS	CHNOS	CHNOS
CHClNO	CHN	CHNS	CHClNO	CHClNO	CHBrO	CHClNO	CHClNO	CHN	CHClNO	CHClNO	CHClNO
CHOS	CHBrO	CHNOS	CHN	CHN	CHNOS	CHBrNO	CHBrO	CHClNO	CHN	CHN	CHOS
CHNOS	CHBrNO	CHClNO	CHBrNO	CHBrNO	CHClO	CHN	CHN	CHClO	CHOS	CHClNOS	CHN
CHN	CHOS	CHOS	CHBrO	CHBrO	CHClNO	CHClO	CHClO	CHOS	CHClO	CHOS	CHClNOS
CHBrO	CHBrN	CHBrO	CHClO	CHClO	CHBrNO	CHBrO	CHBrNO	CHBrNO	CHClNOS	CHClO	CHClO
CHClNPt	CH	CHClO	CHClN	CHClN	CHOS	CHOS	CHOS	CHBrO	CHNS	CHNS	CHBrNO
CH	CHClNO	CHBrNO	CHNS	CHNS	CHINO	CH	CH	CHClN	CH	CHFNO	CHFNO
CHClNPt	CHNaO	CHClNPt	CHCoNOS	ClCoHNO	CHNiOS	CHCuNO	CHClCoNO	CHAgN	CHNNiO	CHOSn	CHCoNO
CHClPPt	BaHOPW	CHAgNO	CHClHgO	CHCoNO	CHHgO	CHNNiO	CHCoNO	CHNNiO	CHCuNO	CHCoNO	CHCuNO
CHNOPt	HKOPW	CHNaOS	CHKNO	BrCoHNO	CHITe	CHClHgNO	CHNNiO	CHCuNO	CHCrO	CHSn	CHFeO
CHKOS	CHCuO	BeNaOSi	CHClNOPt	CHBrCoNO	CHFeNO	CHCoNO	CHNOPtS	CHCoNO	CHOSn	CHCrNO	CHNNiO
CHAsClPt	ClMoO	CHNNaO	CHClNPt	CoHNO	HMoNO	CHNaOS	CHCuNOS	CHHgNO	CHSn	CHOSSn	CHClCoNO
HNaOU	BClKPt	CHAgNOS	CHFeNNaO	CHCrNO	CHKNOS	CHAgNOS	CHBrCoNO	CHAg	CHClNOV	CHNOSn	CHFeNO
BrHNPt	CHAgNO	CHNNaOS	CHNNaO	ClHNTh	CHIOTe	CHAgNO	CHCuNO	CHClHgO	CHAgNO	CHCuNO	CHClCuNO
CHClNOPt	CHBrHgN	CHCrNOS	CHHgO	CoHNOS	HHgNOS	CHCuNOS	CHClNNiO	CHHgO	CHClNNiO	CHCrO	CHNNaOS
CHINPt	CHOPb	AsHNaOS	CHHgNO	CHClHgO	CHIrNO	CHNNiOS	CHClNNi	CHOSn	CHCrNO	CHClOSn	CHNNaO
CHNOPtS	BBaHO	CHClCoNPt	CHAlO	CHClCoO	CHSn	CHHgIS	CHHgIP	CHClCoNO	CHClNV	CHFeNO	CHClNNiO

Table S4. List of the 10 most frequent compositions of elements in some particular years of the third regime. The second half of the table lists the remaining 10 most frequent compositions including metals.

1990	2000	2010	2015
CHNO	CHNO	CHNO	CHNO
CHO	CHO	CHNOS	CHFNO
CHNOS	CHNOS	CHFNO	CHNOS
CHClNO	CHClNO	CHClNO	CHO
CHOS	CHFNO	CHO	CHClNO
CHFNO	CHOSi	CHFNOS	CHFNOS
CHN	CHOS	CHClNOS	CHN
CHOSi	CHClNOS	CHClFNO	CHClFNO
CHClNOS	CHN	CHBrNO	CHBrNO
CHBrNO	CHNOSi	CHN	CHClNOS
CHCuNO	CHFeNO	CHNOZn	CHNOZn
CHFeNO	CHNOZn	CHCuNO	CHCuNO
CHCoNO	CHCuNO	CHNNaOS	CHFeNO
CHFeO	CHNNiO	CHCoNO	CHCoNO
CHNNaOS	CHClCuNO	CHClCuNO	CHNNaOS
CHClCuNO	CHCoNO	CHFeNO	CHIrN
CHNNiO	CHNNaOS	CHNNiO	CHNNiO
CHNORu	CHClNORu	CHNNaO	CHClCuNO
CHClCoNO	CHClFeNO	CHNOPd	CHCdNO
CHOSn	CHFNOPRu	CHCuNOS	CHNOPt

Distance among compositions

For each compound we extracted its elemental composition, which corresponds to the elements present in the compound and that are arranged in lexicographic order. Hence, for H_2SO_4 its composition is HOS. The composition of interest for the chemistry community in year t is given by $c(t) = (w_1(t)c_1(t), w_2(t)c_2(t), \ldots)$, where $w_i(t)$ is the frequency at which the community reports composition $c_i(t)$ in year t. By studying the behavior of $c(t)$ over time, we want to determine whether the community has had several compositional interests, or, on the contrary, its interests have been narrow and focused on few particular compositions. Therefore, we computed the relative frequency $\bar{w}_i(t)$ of composition $c_i(t)$ in year t as:

$$\bar{w}_i(t) = \frac{w_i(t)}{\sum_i w_i(t)}$$

and calculated the composition distance between successive years as:

$$d(t, t+1) = \sum_i |\bar{w}_i(t) - \bar{w}_i(t+1)|.$$

Thus, larger values of d indicate that the community shifts the composition of the reported substances from one year to the other. In contrast, $d = 0$ indicates that compositions of year t remain the same next year.

We calculated d for the time window analyzed (1800-2015) and the resulting plot is shown in Figure S3. It is seen that d decreases to zero as t increases, which shows that $\{\bar{w}_i(t)\}_{t\geq 1}$ behaves like a Cauchy sequence in the composition distribution space. As a result, we expect the existence in the long run of the limits:

$$\bar{w}_i(\infty) = \lim_{t\to\infty} \bar{w}_i(t),$$

i.e. each particular composition has a limit, which is reached when the composition is exhausted.

We are interested in the behavior of d as a function of time, whether it is high or low, increasing or decreasing. A low d indicates that essentially the same compositions are used over the years. This could happen for two opposite reasons:

(1) Chemists try compositions homogeneously, without concentrating on any particular one. This would be an *exploratory* approach, or

(2) they try the same few compositions repeatedly. This would be an extremely *conservative* approach.

Another possibility is that

(3) they try only a few compositions every year, but different ones from year to year. That would be a *semi-conservative* approach. It would result in high values of d.

Our results show that d has dropped over time (Figure S3), indicating that chemists may have populated compositions either homogeneously (1) or conservatively (2). From Figure S4 we infer that chemists have preferred some compositions over others, that is they have followed (2) rather than (1). Moreover, Tables S2 to S4 indicate on which narrow set of compositions chemists have concentrated; in fact, it has been mostly CHNO. In particular, they have definitely not followed strategy (3).

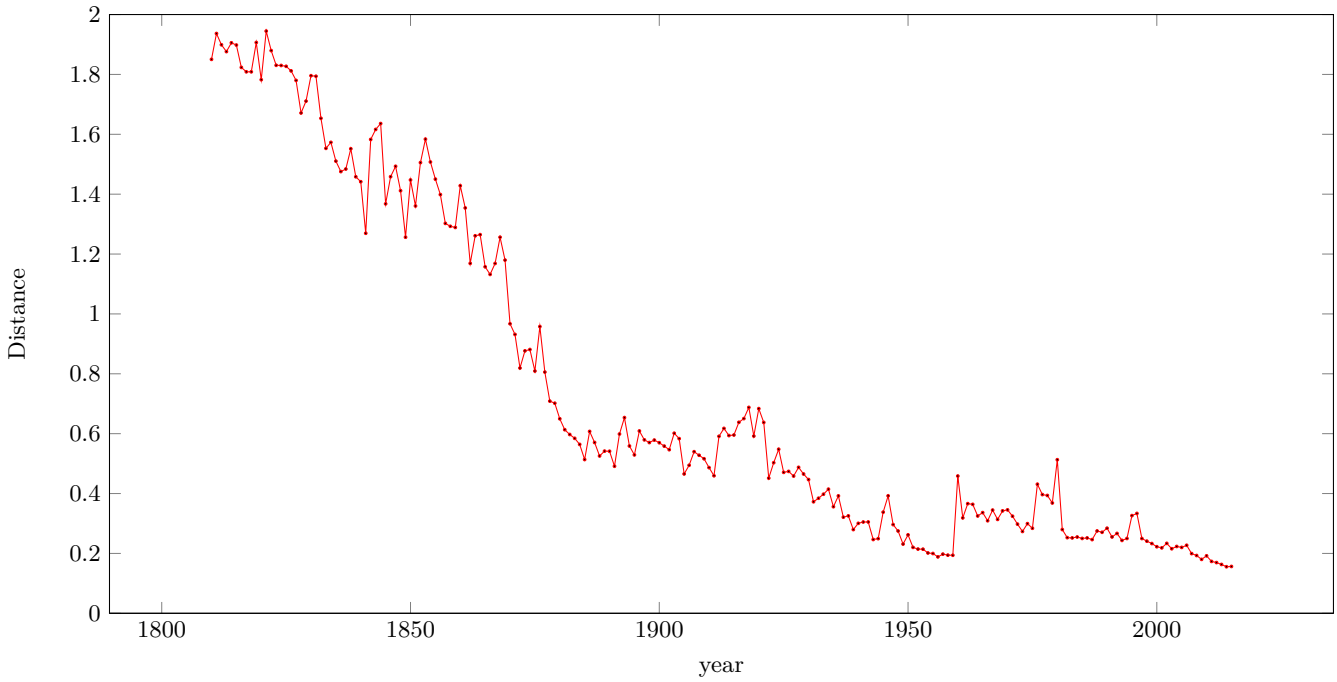

Fig. S3. Distance among elemental compositions of successive years for the period 1800-2015.

Concentration of compositions over time

From each new substance reported in year t, we extracted its composition, which led to a distribution of compositions $c_i(t)$ for t: $c(t) = (w_1(t)c_1(t), w_2(t)c_2(t), \ldots)$ where $w_i(t)$ is the number of new substances matching the composition $c_i(t)$. The plot shows the result by decades, rather than by years; we have simply added the figures for the 10 years of each decade. The box-plot of each decade is shown in (Figure S4) with inter whisker distance accounting for 99.9% of the data.

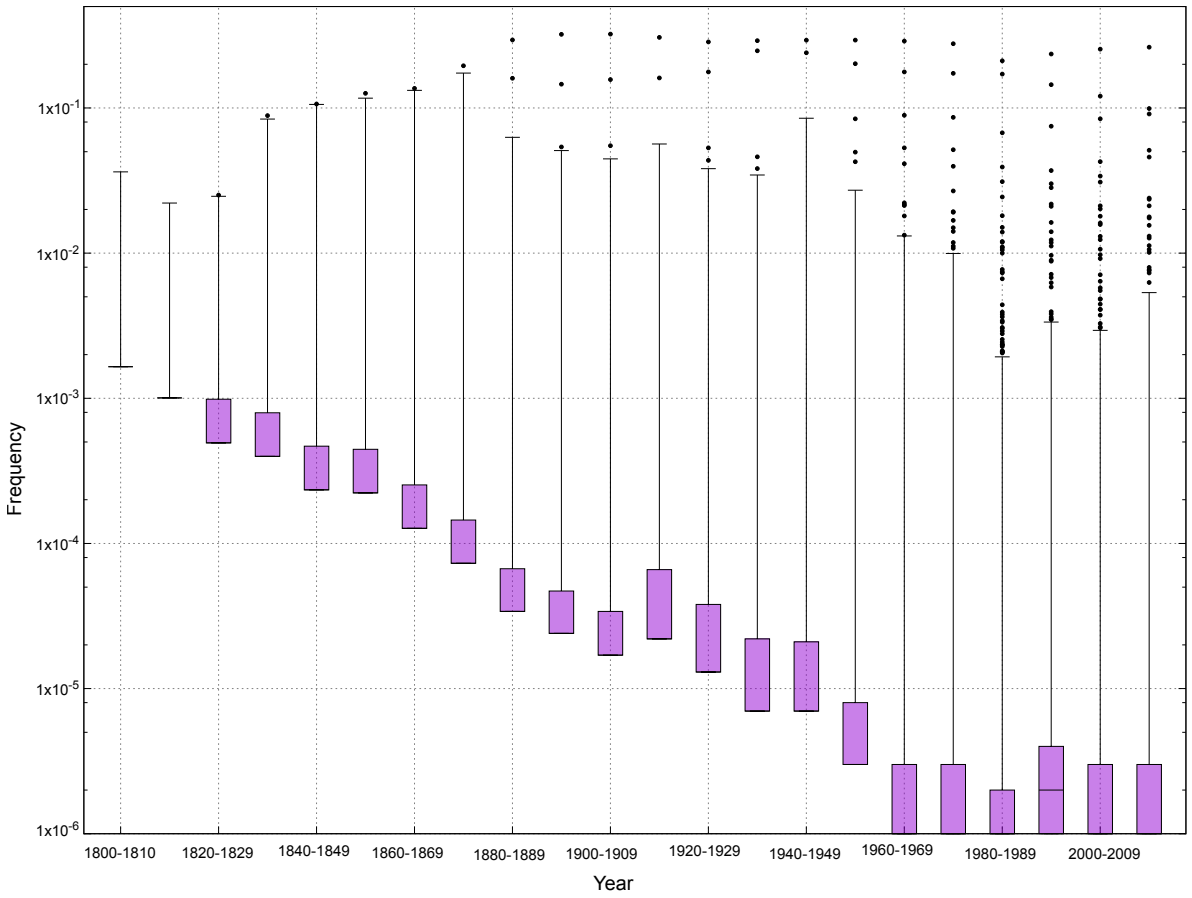

Fig. S4. Box-plots of the compositions every 10 years with inter whisker distance accounting for 99.9% of the data. Relative frequency of compositions are shown at the left axis.

Growth of families of elements

Elements were classified based on the system used in inorganic chemistry textbooks. Here we particularly follow the system by Greenwood and Earnshaw (1) adjusted to classification results based on chemical similarity (2). Classes with more than one element are: Halogens {F, Cl, Br, I}, Noble gases {He, Ne, Ar, Kr, Xe, Rn}, Alkali metals {Li, Na, K, Rb, Cs, Fr}, Alkaline earth metals {Be, Mg, Ca, Sr, Ba, Ra}, {Al, Ga, In, Tl}, {Ge, Sn, Pb}, {As, Sb, Bi}, {Se, Te, Po}, {Ti, Zr, Hf}, {V, Nb, Ta}, {Cr, Mo, W}, {Mn, Tc, Re}, Platinum metals {Fe, Ru, Os, Co, Rh, Ir, Ni, Pd, Pt}, Coinage metals {Cu, Ag, Au}, {Zn, Cd, Hg}, {Ge, Sn, Pb}, Lanthanoids {Sc, Y, La, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu}, Actinoids {Ac, Th, Pa, U, Np, Pu, Am, Cm, Bk, Cf, Es, Fm, Md, No, Lr}.

The annual fraction of new compounds that contain at least one element of each family to the total of new compounds is plotted in Figures S7 to S8.

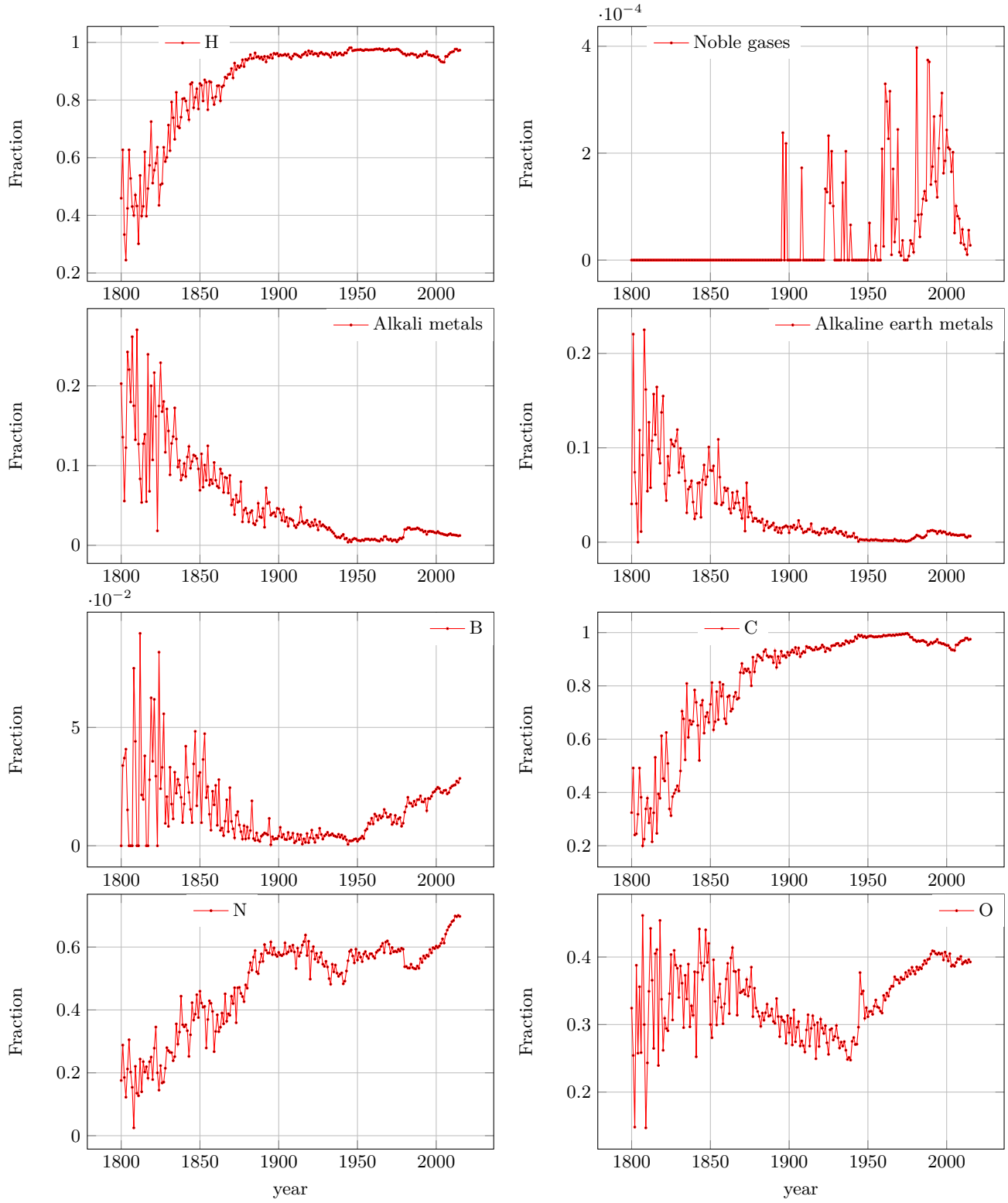

Fig. S5. Fraction of synthesized compounds of the families to the total of new compounds.

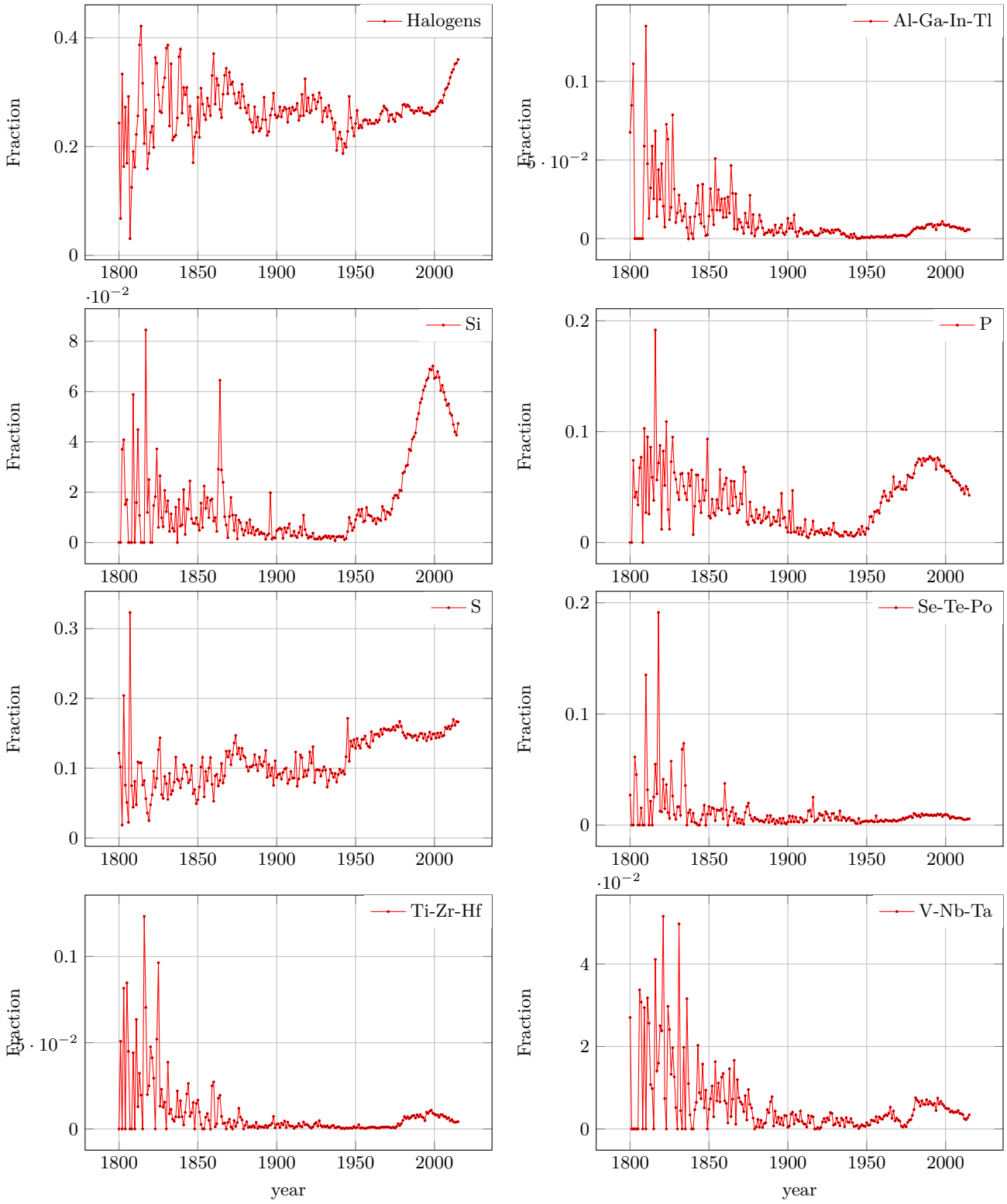

Fig. S6. Fraction of synthesized compounds of the families to the total of new compounds.

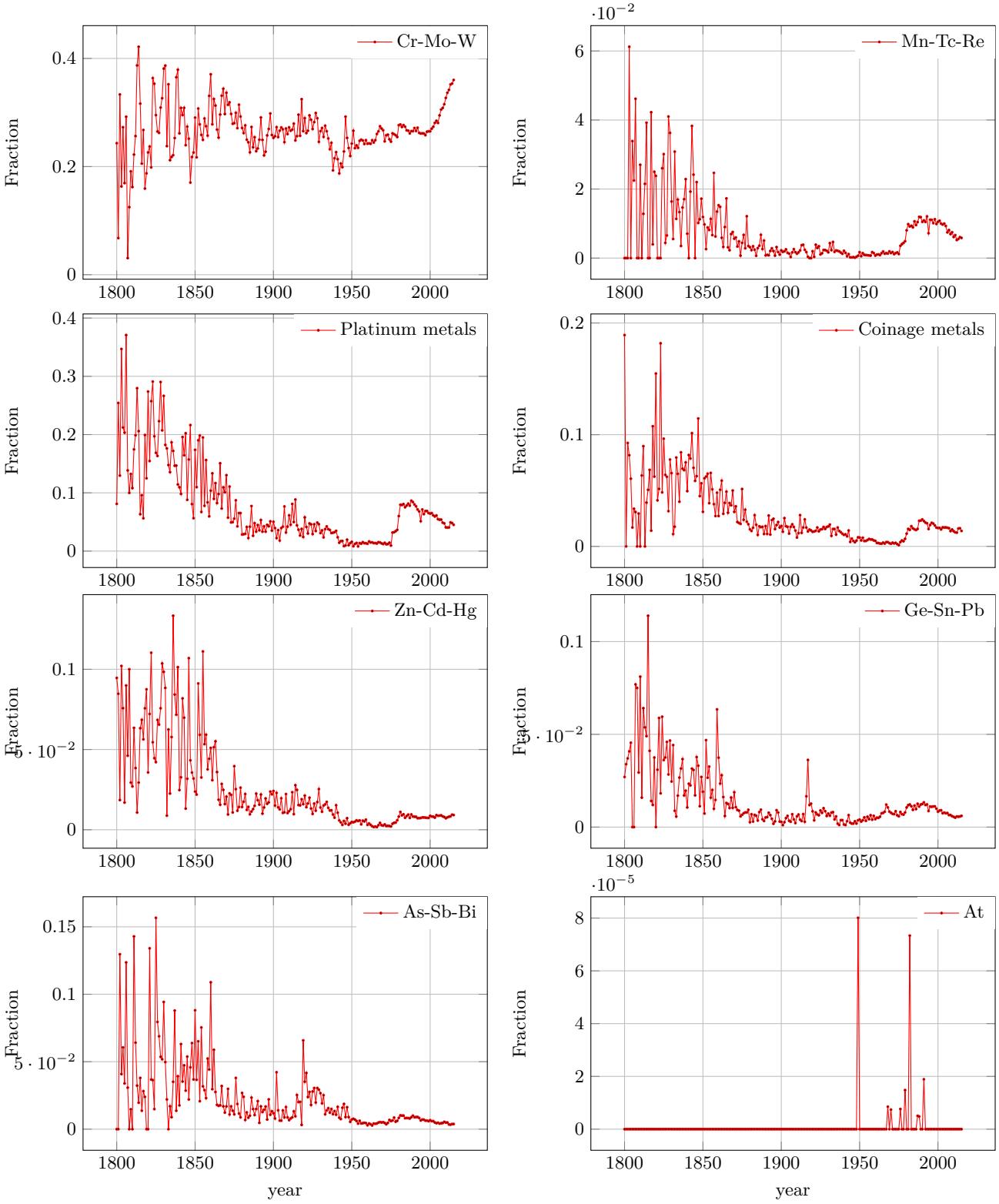

Fig. S7. Fraction of synthesized compounds of the families to the total of new compounds.

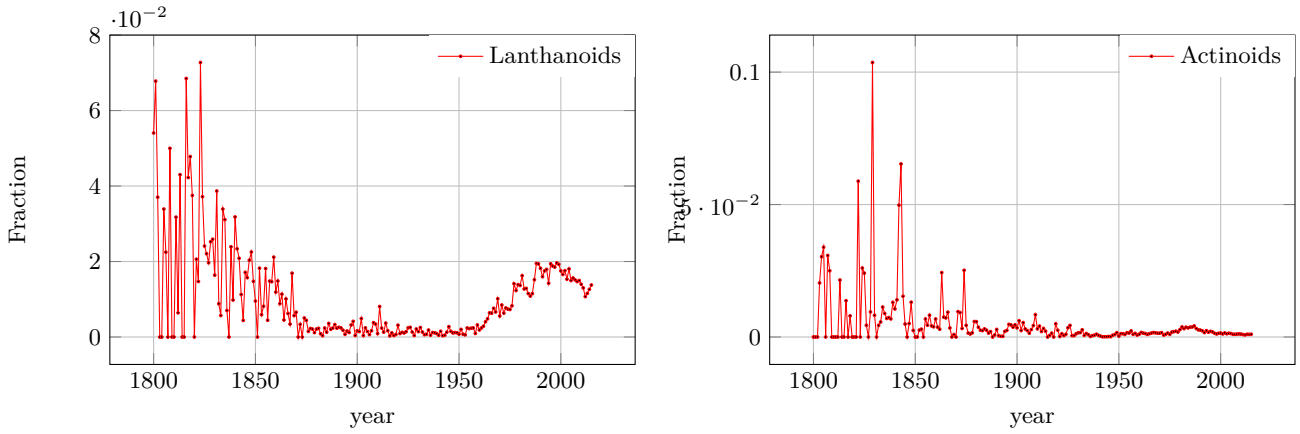

Fig. S8. Fraction of synthesized compounds of the families to the total of new compounds.

Most synthesized products

Table S5 shows the top-10 most frequently synthesized products for each one of the nine periods analyzed (main text).

Table S5. Most synthesized products.

	Before 1860	1860-1879	1880-1889	1900-1919	1920-1939	1940-1959	1960-1979	1980-1999	2000-2015
1	NH_3	NH_3	NH_3	NH_3	BZA	H_2O	H_2O	PhCHO	Glc
2	H_2O	CO_2	CO_2	CO_2	NH_3	CO_2	H_2	CO_2	CO_2
3	CO_2	AcOH	MAC	BZA	CO_2	CH_2O	H_2S	Ethene	PhCHO
4	S	HCl	BZA	H_2O	MAC	Methane	O_2	Methane	Ph_2
5	HCl	BZA	H_2O	$PhNH_2$	AcOH	AcOH	CO	BZA	CuO
6	O_2	MAC	$PhNH_2$	AcOH	H_2O	BZA	CO_2	C_6H_6	H_2
7	Cl_2	H_2O	AcOH	OA	OA	Acetone	$B(OH)_3$	PhAc	ZnO
8	Hg	OA	OA	MAC	FA	HCl	Ag	CO	PhAc
9	SO_2	H_2S	HCl	PhCHO	CH_2O	MAC	N_2	Ph_2	BZA
10	H_2	S	EtOH	HCl	HCl	NH_3	FC	Acetone	CO

Abbreviations: AcOH (Acetic acid), Ag (Silver), $B(OH)_3$ (Boric acid), BZA (Benzoic acid), C_6H_6 (Benzene), CH_2O (Formaldehyde), Cl_2 (Chlorine), CO (Carbon monoxide), CO_2 (Carbon dioxide), CuO (Copper(II) oxide), EDBB (1,1'-(1,2-ethanediyl)bisbenzene), EtOH (Ethanol), FA (Formic acid), FC (Ferrocene), Glc (Glucose), H_2 (Hydrogen), H_2O (Water), HCl (Hydrochloric acid), Hg (Mercury), H_2S (Hydrogen sulfide), MAC (Methylammonium carbonate), N_2 (Nitrogen), NH_3 (Ammonia), O_2 (Oxygen), OA (Oxalic acid), Ph_2 (Biphenyl), PhAc (Acetophenone), PhCHO (Benzaldehyde), $PhNH_2$ (Aniline), S (Sulphur), SO_2 (Sulfur dioxide), ZnO (Zinc oxide).

Analysis of the distribution of substrates

The hypothesis that frequency distributions $p(R)$ of the number of different reactions R in which substrates have participated in period t follow power-law, normal, exponential and Poisson distributions models were tested using the poweRlaw package (3). The mathematical expressions for these distributions are shown in Table S6 (4). To test whether the experimental distribution follows one of the given distributions in Table S6, a goodness-of-fit test was used (3), which generates a p-value that is based on the distance between the experimental distribution and the hypothesized distribution (4). In general, if $p \leq 0.1$, the hypothesized distribution is ruled out. Table S7 shows the results for the distributions of the nine periods and their parameters for substrates.

Table S6. Definition of power-law, normal, exponential and Poisson distributions. For each distribution, the basic functional form $f(R)$ is given along with the normalization constant C such that $\sum_{R=R_{min}}^{\infty} Cf(R) = 1$, where R_{min} is the minimum R value from which the distributions begins to apply. The general expression for these distributions is $p(R) = Cf(R)$.

Name	$f(R)$	C
Power law	$R^{-\alpha}$	$1/\sum_{n=0}^{\infty}(n+R_{min})^{-\alpha}$
Exponential	$e^{-\lambda R}$	$(1-e^{-\lambda})e^{\lambda R_{min}}$
Poisson	$\mu^R/R!$	$[e^{\mu} - \sum_{k=0}^{R_{min}-1} \frac{\mu^k}{k!}]^{-1}$
Log-normal	$\frac{1}{R}\exp[-\frac{(\ln R-\mu)^2}{2\sigma^2}]$	$\sqrt{\frac{2}{\pi\sigma^2}}[\text{erfc}^*(\frac{\ln R_{min}-\mu}{\sqrt{2}\sigma})]^{-1}$

*erfc stands for the complementary error function (5).

Table S7. Parameters and p-values for the distributions shown in Table S6, when applied to the distribution of substrates. The p-values were calculated running 1,000 simulations.

	Parameters			
Period	Power-law	Exponential	Poisson	Log-normal
Before 1860	$R_{min} = 2$ $\alpha = 1.963773$ $p = 0.01$	$R_{min} = 13$ $\lambda = 0.03506312$ $p = 0$	$R_{min} = 13$ $\mu = 41.02849$ $p = 0$	$R_{min} = 2$ $\mu = -5.196269$ $\sigma = 2.827289$ $p = 0.374$
1860-1879	$R_{min} = 1$ $\alpha = 2.016267$ $p = 0$	$R_{min} = 1$ $\lambda = 0.7320951$ $p = 0$	$R_{min} = 771$ $\mu = 1106$ $p = 0.089$	$R_{min} = 1$ $\mu = -6.733486$ $\sigma = 2.985701$ $p = 0.291$
1880-1899	$R_{min} = 9$ $\alpha = 2.113$ $p = 0.49$	$R_{min} = 1$ $\lambda = 0.6836599$ $p = 0$	$R_{min} = 4398$ $\mu = 4815.667$ $p = 0.587$	$R_{min} = 2$ $\mu = -15.148219$ $\sigma = 4.012288$ $p = 0$
1900-1919	$R_{min} = 3$ $\alpha = 2.177336$ $p = 0$	$R_{min} = 1$ $\lambda = 0.6561324$ $p = 0$	$R_{min} = 3839$ $\mu = 6352.571$ $p = 0.038$	$R_{min} = 18$ $\mu = -50.422936$ $\sigma = 7.343984$ $p = 0.587$
1920-1939	$R_{min} = 2$ $\alpha = 2.160733$ $p = 0$	$R_{min} = 1$ $\lambda = 0.6735098$ $p = 0$	$R_{min} = 4177$ $\mu = 8774.154$ $p = 0$	$R_{min} = 27$ $\mu = -19.421922$ $\sigma = 4.956338$ $p = 0.837$
1940-1959	$R_{min} = 18$ $\alpha = 2.011466$ $p = 0.11$	$R_{min} = 1$ $\lambda = 0.7368087$ $p = 0$	$R_{min} = 8638$ $\mu = 13026.17$ $p = 0.029$	$R_{min} = 23$ $\mu = -15.601899$ $\sigma = 4.544417$ $p = 0.916$
1960-1979	$R_{min} = 1$ $\alpha = 2.340214$ $p = 0$	$R_{min} = 2$ $\lambda = 0.8119482$ $p = 0$	$R_{min} = 3574$ $\mu = 7080.636$ $p = 0.001$	$R_{min} = 50$ $\mu = -0.3703384$ $\sigma = 2.3445863$ $p = 0.927$
1980-1999	$R_{min} = 1$ $\alpha = 2.341685$ $p = 0$	$R_{min} = 1$ $\mu = 0.8174186$ $p = 0$	$R_{min} = 23031$ $\mu = 41872.17$ $p = 0.064$	$R_{min} = 94$ $\mu = -3.513336$ $\sigma = 3.218282$ $p = 0.101$
2000-2015	$R_{min} = 5$ $\alpha = 2.114837$ $p = 0$	$R_{min} = 2$ $\mu = 0.8376273$ $p = 0$	$R_{min} = 46242$ $\mu = 61491.5$ $p = 0.052$	$R_{min} = 1$ $\mu = -1035.13808$ $\sigma = 28.84289$ $p = 0.064$

Analysis of jumps for substrates

The distribution of participation of compounds in R different reactions as substrates for the following periods is shown in Figure 2a (main text): Period 1 (Before 1860), Period 2 (1860-1879), Period 3 (1880-1899), Period 4 (1900-1919), Period 5 (1920-1939), Period 6 (1940-1959), Period 7 (1960-1979), Period 8 (1980-1999), Period 9 (2000-2015).

We found that the participation of substrates in less than 21 reactions ($R \leq 20$) accounts for 98% of the whole variance of the distribution (Figure 2a, main text). Therefore, we plotted (Figure S9) the logarithm growth of participation of compounds in $R \leq 20$ different reactions as substrates in the nine periods of Figure 2a. The least squares regression method shows that the logarithm growth of participation of compounds $\log \text{Comp}_i$, with $i = 1, \ldots, 9$ follows the equation

$$\log \text{Comp}_i = 0.7839528i + 7.9552287 + \text{Residual}_i,$$

where $\text{Residual}_1 = 0.174906828$, $\text{Residual}_2 = -0.084145563$, $\text{Residual}_3 = 0.221563190$, $\text{Residual}_4 = -0.107935989$, $\text{Residual}_5 = -0.237065832$, $\text{Residual}_6 = -0.300628224$, $\text{Residual}_7 = -0.073863858$, $\text{Residual}_8 = 0.397940841$, and $\text{Residual}_9 = 0.009228608$.

These residuals pass the Shapiro-Wilk and Kolmogorov-Smirnov normality tests for mean 0 and standard deviation $sd = 0.2260776$, with the corresponding probabilities $p^{ShW} = 0.7444$ and $p^{KS} = 0.8711$. There are two historical jumps, namely between 1860-1879 and 1880-1899, and between 1960-1979 and 1980-1999, corresponding to $\text{Residual}_3 = 0.221563190$ and $\text{Residual}_8 = 0.397940841$, which are higher than the other residuals, and still in the range $(-2sd, 2sd)$. As a result, the logarithm difference $\log \text{Comp}_i - \log \text{Comp}_{i-1}$, with $i = 2, \ldots 9$ reaches the highest and the second highest at these two periods.

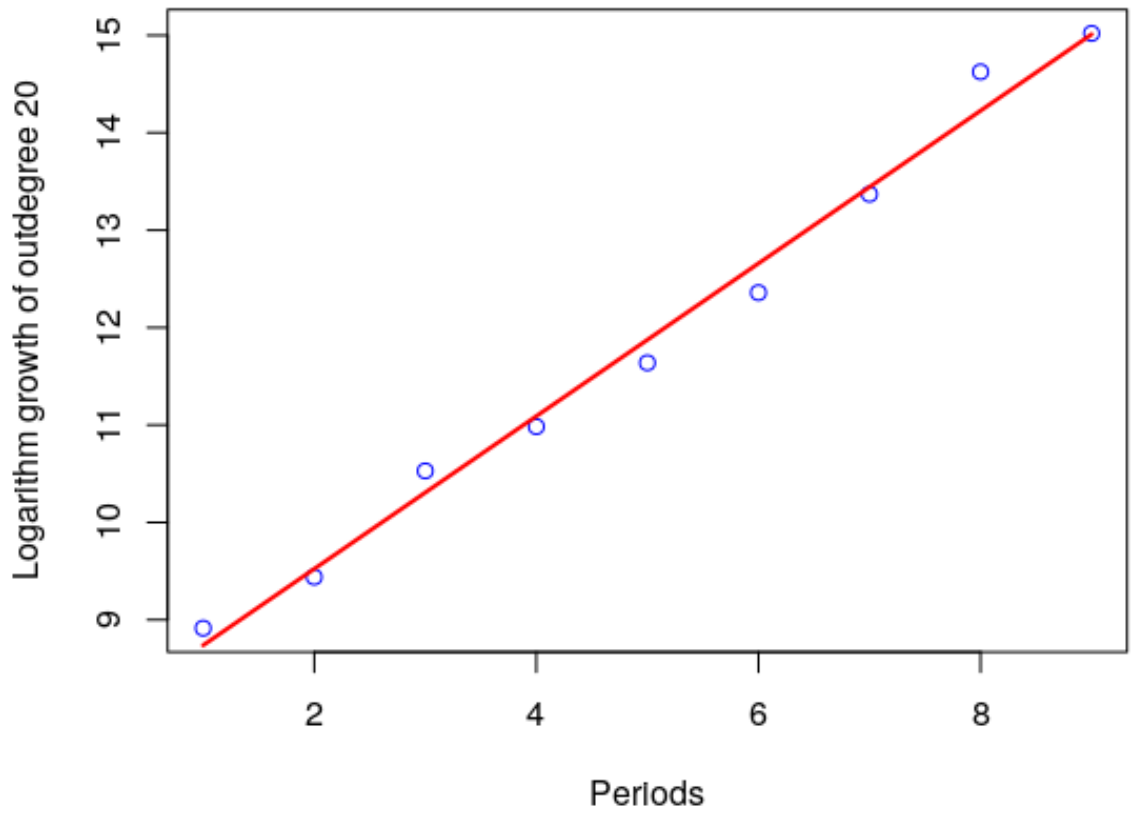

Fig. S9. Logarithm growth of participation of compounds in $R \leq 20$ different reactions as substrates in the nine periods.

Analysis of the distribution of products

The hypothesis that frequency distributions $p(R)$ of the number of different reactions R in which products have participated in period t follow power-law, normal, exponential and Poisson distributions models were tested using the the poweRlaw package (3). The mathematical expressions for these distributions are shown in Table S6 (4). To test whether the experimental

distribution follows one of the given distributions in Table S6, a goodness-of-fit test was used (3), which generates a p-value that is based on the distance between the experimental distribution and the hypothesized distribution (4). In general, if $p \leq 0.1$, the hypothesized distribution is ruled out. Table S8 shows the results for the distributions of the nine periods and their parameters for products.

Table S8. Parameters and p-values for the distributions shown in Table S6, when applied to the distribution of products. The p-values were calculated running 1,000 simulations.

	Parameters			
Period	Power-law	Exponential	Poisson	Log-normal
Before 1860	$R_{min} = 2$ $\alpha = 2.314181$ $p = 0.049$	$R_{min} = 59$ $\mu = 0.0187567$ $p = 0.718$	$R_{min} = 161$ $\mu = 203.6254$ $p = 0.207$	$R_{min} = 1$ $\mu = -718.58135$ $\sigma = 23.31244$ $p = 0.049$
1860-1879	$R_{min} = 1$ $\alpha = 2.451058$ $p = 0$	$R_{min} = 68$ $\mu = 0.01125429$ $p = 0.941$	$R_{min} = 1$ $\mu = 1.671426$ $p = 0$	$R_{min} = 1$ $\mu = -16.007938$ $\sigma = 3.667497$ $p = 0.011$
1880-1899	$R_{min} = 5$ $\alpha = 2.479312$ $p = 0.513$	$R_{min} = 2$ $\mu = 0.8289619$ $p = 0$	$R_{min} = 1$ $\mu = 1.511293$ $p = 0$	$R_{min} = 1$ $\mu = -3.909269$ $\sigma = 1.944246$ $p = 0$
1900-1919	$R_{min} = 8$ $\alpha = 2.467412$ $p = 0.476$	$R_{min} = 2$ $\mu = 0.8198701$ $p = 0$	$R_{min} = 1$ $\mu = 1.435699$ $p = 0$	$R_{min} = 1$ $\mu = -1.929820$ $\sigma = 1.488447$ $p = 0$
1920-1939	$R_{min} = 11$ $\alpha = 2.388553$ $p = 0.881$	$R_{min} = 2$ $\mu = 0.8359276$ $p = 0$	$R_{min} = 1$ $\mu = 1.333986$ $p = 0$	$R_{min} = 20$ $\mu = -17.738687$ $\sigma = 4.029873$ $p = 0.881$
1940-1959	$R_{min} = 1$ $\alpha = 2.947494$ $p = 0$	$R_{min} = 2$ $\mu = 1.148441$ $p = 0$	$R_{min} = 1$ $\mu = 0.7163911$ $p = 0$	$R_{min} = 1$ $\mu = -3.183353$ $\sigma = 1.522464$ $p = 0$
1960-1979	$R_{min} = 1$ $\alpha = 4.126991$ $p = 0$	$R_{min} = 63$ $\mu = 0.0152121$ $p = 0.13$	$R_{min} = 1$ $\mu = 0.2046896$ $p = 0$	$R_{min} = 1$ $\mu = -10.352775$ $\sigma = 2.075666$ $p = 0$
1980-1999	$R_{min} = 1$ $\alpha = 3.211598$ $p = 0$	$R_{min} = 2$ $\mu = 0.9182153$ $p = 0$	$R_{min} = 1$ $\mu = 0.576151$ $p = 0$	$R_{min} = 1$ $\mu = -8.801885$ $\sigma = 2.259969$ $p = 0$
2000-2015	$R_{min} = 1$ $\alpha = 3.672836$ $p = 0$	$R_{min} = 2$ $\mu = 0.9465415$ $p = 0$	$R_{min} = 1$ $\mu = 0.3613914$ $p = 0$	$R_{min} = 1$ $\mu = -206.291472$ $\sigma = 9.765858$ $p = 0$

Similarity of distribution of products

The distance between frequency distributions of participation of compounds in R different reactions as products of Figure 2b (main text) is given by:

$$d(d_i, d_j) = \frac{\sum_{k=1}^{20} |f_i(R_k) - f_j(R_k)|}{f_i(R_k) + f_j(R_k)},$$

where d_l is a frequency distribution and $f_l(R_l)$ is the frequency of participation of products in R_l reactions. Note that only products obtained in less than 21 reactions ($R \leq 20$) where considered, for they account for 97% of the whole variance of the distributions (Figure 2b, main text). The results of the pair-wise distances for the nine periods shown in Figure 2b are found in Table S9. It is observed that the closest distributions are d_8 and d_9, which correspond to periods 1980-1999 and 2000-2015, respectively.

Table S9. Distances between distributions of Figure 2b (main text).

	1860-1879	1880-1899	1900-1919	1920-1939	1940-1959	1960-1979	1980-1999	2000-2015
Before 1860	6.00659	13.8973	15.4195	17.3573	17.4216	15.2836	19.5405	19.5062
1860-1879		10.0815	12.4142	15.4533	15.6046	12.5027	19.1822	19.1245
1880-1899			3.54215	8.94705	9.52587	5.48084	17.7134	17.5695
1900-1919				5.90002	6.70359	4.84538	16.87	16.688
1920-1939					2.06348	7.05896	14.6438	14.3598
1940-1959						6.10315	14.2749	14.0378
1960-1979							15.5765	15.6039
1980-1999								1.00596

Analysis of jumps for products

The distribution of participation of compounds in R different reactions as products for the following periods is shown in Figure 2b (main text): Period 1 (Before 1860), Period 2 (1860-1879), Period 3 (1880-1899), Period 4 (1900-1919), Period 5 (1920-1939), Period 6 (1940-1959), Period 7 (1960-1979), Period 8 (1980-1999), Period 9 (2000-2015).

We found that the participation of products in less than 21 reactions ($R \leq 20$) accounts for 97% of the whole variance of the distribution (Figure 2b, main text). Therefore, we plotted (Figure S10) the logarithm growth of participation of compounds in $R \leq 20$ different reactions as products in the nine periods of Figure 2b. The least squares regression method shows that the logarithm growth of participation of compounds $\log \mathrm{Comp}_i$, with $i = 1, \ldots, 9$ follows the equation

$$\log \mathrm{Comp}_i = 0.8447904i + 8.3520785 + \mathrm{Residual}_i,$$

where $\mathrm{Residual}_1 = 0.04787297$, $\mathrm{Residual}_2 = -0.17878533$, $\mathrm{Residual}_3 = 0.28512578$, $\mathrm{Residual}_4 = -0.07595590$, $\mathrm{Residual}_5 = -0.20859164$, $\mathrm{Residual}_6 = -0.16903558$, $\mathrm{Residual}_7 = 0.35814213$, $\mathrm{Residual}_8 = 0.16272740$, and $\mathrm{Residual}_9 = -0.22149983$.

These residuals pass the Shapiro-Wilk and Kolmogorov-Smirnov normality tests for mean 0 and standard deviation $sd = 0.2228266$, with the corresponding probabilities $p^{ShW} = 0.1398$ and $p^{KS} = 0.6966$. There are two historical jumps, namely between 1860-1879 and 1880-1899, and between 1940-1959 and 1960-1979, corresponding to $\mathrm{Residual}_3 = 0.28512578$ and $\mathrm{Residual}_7 = 0.35814213$, which are higher than the standard deviation $sd = 0.2228266$ (and also higher than the other residuals), and still in the range $(-2sd, 2sd)$. As a result, the logarithm difference $\log \mathrm{Comp}_i - \log \mathrm{Comp}_{i-1}$, with $i = 2, \ldots 9$ reaches the highest and the second highest at these two periods.

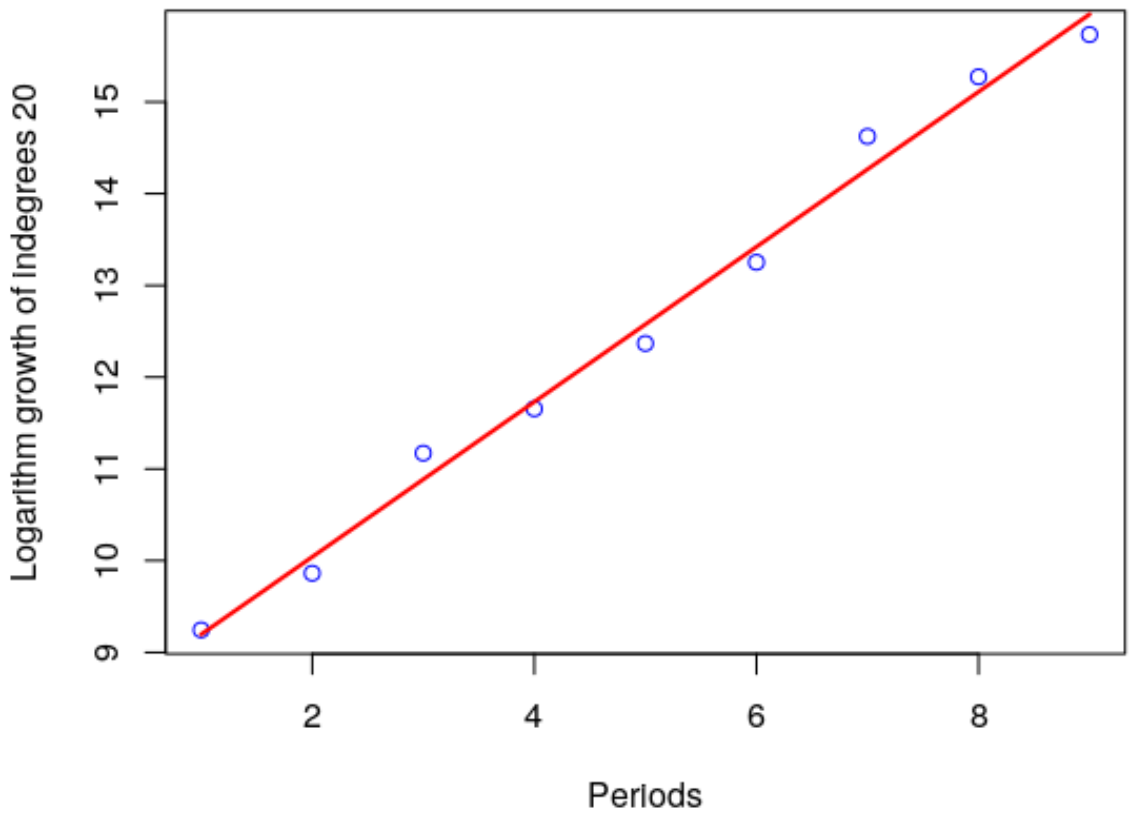

Fig. S10. Logarithm growth of participation of compounds in $R \leq 20$ different reactions as products in the nine periods.

Analysis of the distribution of targets

The frequency distributions of the number of reactions R producing the same target is shown in Figure S11.

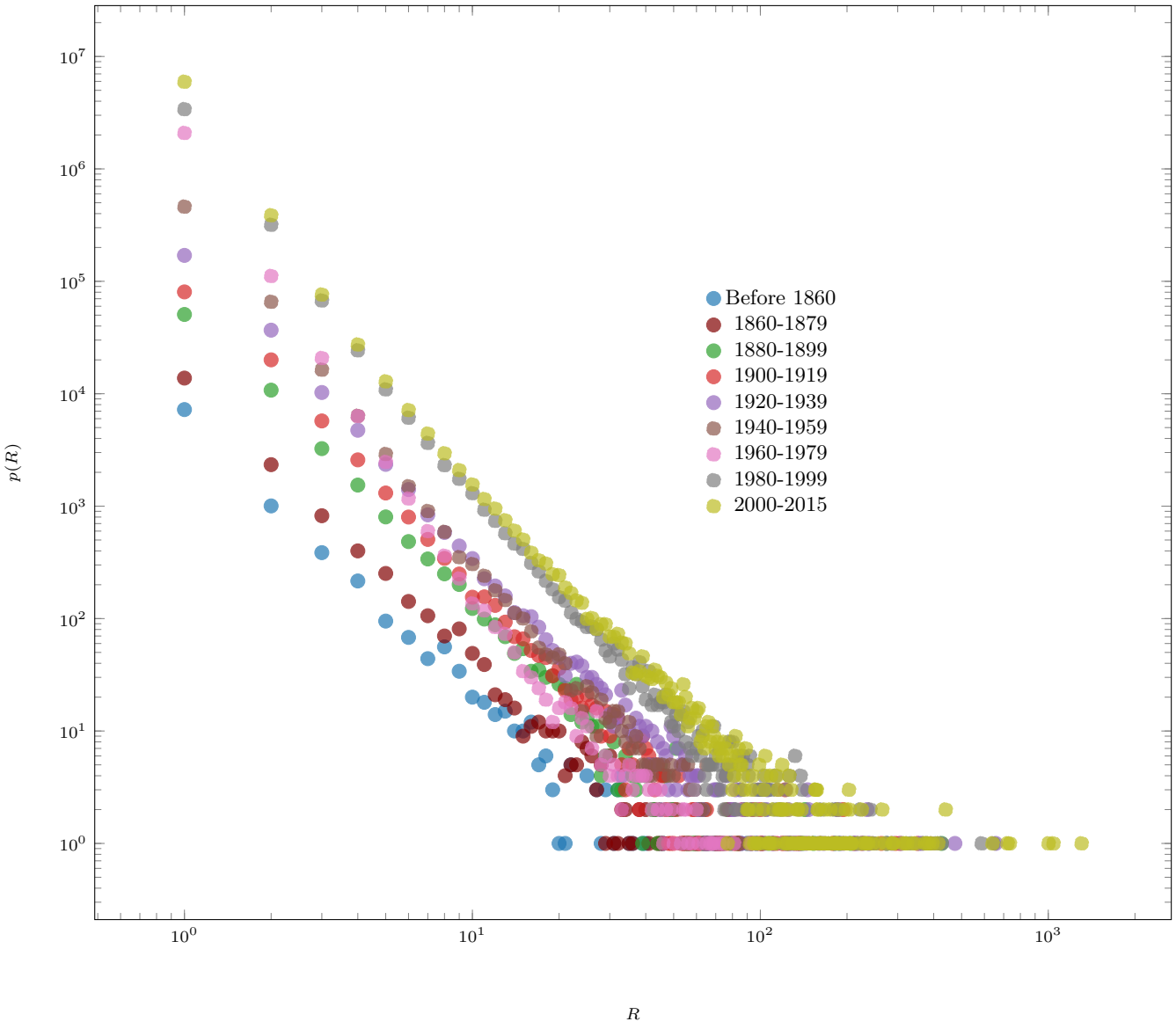

Fig. S11. Frequency distributions of number of reactions R producing the same target.

The hypothesis that frequency distributions $p(R)$ of the number of different reactions R in which targets have participated in period t follow power-law, normal, exponential and Poisson distributions models were tested using the poweRlaw package (3). The mathematical expressions for these distributions are shown in Table S6 (4). To test whether the experimental distribution follows one of the given distributions in Table S6, a goodness-of-fit test was used (3), which generates a p-value that is based on the distance between the experimental distribution and the hypothesized distribution (4). In general, if $p \leq 0.1$, the hypothesized distribution is ruled out. Table S10 shows the results for the distributions of the nine periods and their parameters for targets.

Table S10. Parameters and p-values for the distributions shown in Table S6, when applied to the distribution of targets. The p-values were calculated running 1,000 simulations.

	Parameters			
Period	Power-law	Exponential	Poisson	Log-normal
Before 1860	$R_{min} = 2$ $\alpha = 2.417702$ $p = 0.186$	$R_{min} = 17$ $\mu = 0.06030108$ $p = 0.171$	$R_{min} = 56$ $\mu = 63.30087$ $p = 0.721$	$R_{min} = 1$ $\mu = -30.626118$ $\sigma = 4.792057$ $p = 0.186$
1860-1879	$R_{min} = 1$ $\alpha = 2.532562$ $p = 0.189$	$R_{min} = 12$ $\mu = 0.08234484$ $p = 0.003$	$R_{min} = 1$ $\mu = 1.269463$ $p = 0$	$R_{min} = 1$ $\mu = -7.279544$ $\sigma = 2.480881$ $p = 0.197$
1880-1899	$R_{min} = 3$ $\alpha = 2.658096$ $p = 0.223$	$R_{min} = 84$ $\mu = 0.01255286$ $p = 0.67$	$R_{min} = 1$ $\mu = 1.216067$ $p = 0$	$R_{min} = 5$ $\mu = -5.920010$ $\sigma = 2.291455$ $p = 0.584$
1900-1919	$R_{min} = 3$ $\alpha = 2.795804$ $p = 0.001$	$R_{min} = 47$ $\mu = 0.01875625$ $p = 0.064$	$R_{min} = 1$ $\mu = 1.197611$ $p = 0$	$R_{min} = 7$ $\mu = -20.806110$ $\sigma = 3.752783$ $p = 0.001$
1920-1939	$R_{min} = 7$ $\alpha = 2.636928$ $p = 0.063$	$R_{min} = 90$ $\mu = 0.01074621$ $p = 0.514$	$R_{min} = 1$ $\mu = 1.094216$ $p = 0$	$R_{min} = 11$ $\mu = -11.385738$ $\sigma = 3.071208$ $p = 0.738$
1940-1959	$R_{min} = 2$ $\alpha = 3.339249$ $p = 0$	$R_{min} = 68$ $\mu = 0.0181202$ $p = 0.886$	$R_{min} = 1$ $\mu = 0.5954598$ $p = 0$	$R_{min} = 1$ $\mu = -2.467482$ $\sigma = 1.333142$ $p = 0$
1960-1979	$R_{min} = 1$ $\alpha = 4.198562$ $p = 0$	$R_{min} = 1$ $\mu = 2.454067$ $p = 0$	$R_{min} = 1$ $\mu = 0.1825004$ $p = 0$	$R_{min} = 1$ $\mu = -6.231822$ $\sigma = 1.609130$ $p = 0$
1980-1999	$R_{min} = 1$ $\alpha = 3.479212$ $p = 0$	$R_{min} = 2$ $\mu = 1.154287$ $p = 0$	$R_{min} = 1$ $\mu = 0.390401$ $p = 0$	$R_{min} = 1$ $\mu = -6.062233$ $\sigma = 1.795392$ $p = 0$
2000-2015	$R_{min} = 1$ $\alpha = 3.855568$ $p = 0$	$R_{min} = 2$ $\mu = 0.954044$ $p = 0$	$R_{min} = 1$ $\mu = 0.2859205$ $p = 0$	$R_{min} = 1$ $\mu = -118.983190$ $\sigma = 7.210826$ $p = 0$

References

1. Greenwood NN, Earnshaw A (1998) *Chemistry of the elements*. (Butterworth Heinemann).
2. Leal W, Restrepo G, Bernal A (2012) A network study of chemical elements: From binary compounds to chemical trends. *MATCH Commun. Math. Comput. Chem.* 68:417–442.
3. Gillespie C (2015) Fitting heavy tailed distributions: The powerlaw package. *Journal of Statistical Software, Articles* 64(2):1–16.
4. Clauset A, Shalizi C, Newman M (2009) Power-law distributions in empirical data. *SIAM Review* 51(4):661–703.
5. Gautschi W (1972) *Error function and Fresnel integrals*, eds. Abramowitz M, Stegun IA. (Dover), p. 299.